

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:37:07 ; Search time 470.56 Seconds
(without alignments)
33.359 Million cell updates/sec

Title: US-09-396-196f-4

Sequence: 1 ggalctgtcttgaagcagcagcag 25

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq.0601.*
1: /SIDSI/gcgcdata/geneseq/geneseqn/NA1980.DAT.*
2: /SIDSI/gcgcdata/geneseq/geneseqn/NA1981.DAT.*
3: /SIDSI/gcgcdata/geneseq/geneseqn/NA1982.DAT.*
4: /SIDSI/gcgcdata/geneseq/geneseqn/NA1983.DAT.*
5: /SIDSI/gcgcdata/geneseq/geneseqn/NA1984.DAT.*
6: /SIDSI/gcgcdata/geneseq/geneseqn/NA1985.DAT.*
7: /SIDSI/gcgcdata/geneseq/geneseqn/NA1986.DAT.*
8: /SIDSI/gcgcdata/geneseq/geneseqn/NA1987.DAT.*
9: /SIDSI/gcgcdata/geneseq/geneseqn/NA1988.DAT.*
10: /SIDSI/gcgcdata/geneseq/geneseqn/NA1989.DAT.*
11: /SIDSI/gcgcdata/geneseq/geneseqn/NA1990.DAT.*
12: /SIDSI/gcgcdata/geneseq/geneseqn/NA1991.DAT.*
13: /SIDSI/gcgcdata/geneseq/geneseqn/NA1992.DAT.*
14: /SIDSI/gcgcdata/geneseq/geneseqn/NA1993.DAT.*
15: /SIDSI/gcgcdata/geneseq/geneseqn/NA1994.DAT.*
16: /SIDSI/gcgcdata/geneseq/geneseqn/NA1995.DAT.*
17: /SIDSI/gcgcdata/geneseq/geneseqn/NA1996.DAT.*
18: /SIDSI/gcgcdata/geneseq/geneseqn/NA1997.DAT.*
19: /SIDSI/gcgcdata/geneseq/geneseqn/NA1998.DAT.*
20: /SIDSI/gcgcdata/geneseq/geneseqn/NA1999.DAT.*
21: /SIDSI/gcgcdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDSI/gcgcdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	20	AAO1303
2	25	100.0	1084	10	AAO1329
3	25	100.0	1121	7	AAO60496
4	25	100.0	5872	15	AAO62386
5	18.6	74.4	2515	21	AAO92058
6	18.6	74.4	8867	21	AAO92058
7	17.6	70.4	808	20	AAO20711
8	17.6	70.4	3561	16	AAO85414
9	17.6	70.4	3561	21	AAO6246
10	17.6	70.4	15894	14	AAO40480
11	17.6	70.4	15894	19	AAO18264

12	17.6	70.4	15894	19	AAO18268
13	17.6	70.4	15894	19	AAO18269
14	17.6	70.4	15894	19	AAO18270
15	17.6	70.4	15894	19	AAO18271
16	17.6	70.4	15894	19	AAO18271
17	17.6	70.4	15894	19	AAO18265
18	17.6	70.4	15894	19	AAO18265
19	17.6	70.4	15894	20	AAO22902
20	17.6	70.4	15894	20	AAO22903
21	17.6	70.4	15894	20	AAO22905
22	17.6	70.4	15894	20	AAO22907
23	17.6	70.4	15894	20	AAO22908
24	17.6	70.4	15894	20	AAO22906
25	17.4	69.6	801	21	AAO59384
26	17.2	68.8	1970	21	AAO16338
27	17.2	68.8	2126	21	AAO44857
28	17.2	68.8	2203	21	AAO256959
29	17.2	68.8	5504	20	AAO20559
30	17	68.0	252	21	AAO61129
31	17	68.0	312	21	AAO75304
32	17	68.0	394	21	AAO7943
33	17	68.0	1579	21	AAO93432
34	17	68.0	1711	21	AAO76996
35	17	68.0	2040	21	AAO61127
36	17	68.0	2153	9	AAO80922
37	17	68.0	2526	20	AAO6819
38	17	68.0	2526	21	AAO28691
39	17	68.0	2612	21	AAO61153
40	17	68.0	2750	21	AAO28690
41	17	68.0	2895	21	AAO46633
42	17	68.0	3619	21	AAO61126
43	17	68.0	1230025	20	AAO91990
44	16.8	67.2	365	21	AAO6234
45	16.8	67.2	404	21	AAO6233

ALIGNMENTS

RESULT 1	
AAO1303	standard; DNA; 1041 BP.
ID	AAO1303
AC	AAO1303
XX	
XX	12-APR-1999 (first entry)
DE	E. coli biotin synthetase (Biot) coding sequence.
XX	
KW	DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW	biotin synthetase; biotin production; vitamin H; Biot; ss.
OS	Escherichia coli.
XX	
PN	US5869719-A.
XX	
PD	09-FEB-1999.
XX	
PE	30-APR-1997; 97US-0846338.
XX	
PR	30-APR-1997; 97US-0846338.
PR	08-MAR-1995; 95US-0401068.
XX	
PA	(NOVS) NOVARTIS FINANCE CORP.
XX	
PI	Patton DA;
XX	
DR	WPI; 1999-152902/13.
XX	
XX	P-PSDB; AAO73906.
XX	
PT	Transgenic plants with high biotin levels - transformed with DNA
PT	encoding di-aminopelargonic acid amino-transferase or biotin
PT	synthase

XX Example 2; Column 37-40; 34pp; English.

CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.

SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.038;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgacag 25
 ||||||||||||||||||
 Db 63 ggaatcgtctgttgaagcgacag 87

RESULT 2

ID AAN91329 standard; DNA: 1084 BP.

AC AAN91329;

DT 15-FEB-1990 (first entry)

DE E.coli Bio B gene.

KW E.coli; Bio B gene; biotin.

OS Escherichia coli.

FT Key Location/Qualifiers
 CDS 24..1064
 /*tag=a

PN GB2216530-A.

PD 11-OCT-1989.

PE 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

PR 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISH.

PI Pearson BM, McKee RA;

DR WPI; 1989-295085/41. P-PSDB P91392

PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 derived from E.coli and capable of replication and expression in other
 microorganisms, esp. yeast.

PS Table 3; page 33-4; 52pp; English.

CC The gene can be used in a plasmid for expression of enzymes of the biotin
 CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA931, pMA36c, pKY49 and pCK495, and plasmid pCK965 for
 CC lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.

SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.038;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgacag 25
 ||||||||||||||||||
 Db 86 ggaatcgtctgttgaagcgacag 110

RESULT 3

ID AAN60496 standard; DNA: 1121 BP.

AC AAN60496;

DT 17-OCT-1991 (first entry)

DE Sequence encoding biotin synthesising enzyme.

KW Biotin synthetic enzyme; E.coli; desthiobiotin; ds.

FT Key Location/Qualifiers
 CDS 42..1082
 /*tag= a

PN JP61149091-A.

PD 07-JUL-1986.

PE 24-DEC-1984; 84JP-0272605.

PR 24-DEC-1984; 84JP-0272605.

PA (NIPS) NIPPON SODA KK.

DR WPI; 1986-216622/33.

DR P-PSDB; AAP60536.

PT Double stranded DNA encoding biotin synthesising enzyme -
 comprises transformed mutant E.coli strain contg. cyclic doubled
 stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

PS Disclosure; Page 534; 23pp; Japanese.

CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desthiobiotin.

SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.039;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgacag 25
 ||||||||||||||||||
 Db 104 ggaatcgtctgttgaagcgacag 128

RESULT 4

ID AA062386 standard; DNA: 5872 BP.

AC AA062386;

DT 16-NOV-1994 (first entry)

DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

KW Biotin; expression; enterobacteria; vitamin H; synthesis;

KW plasmid; pB030A-15/9; bioB; bioC; bioD; bioA;
 KW promoter; ptac; biotin synthase; KAPA synthase;
 KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;


```

KM dethiobiotin synthase; DAPA synthase;
KM S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
XX seborrhoea; dermatitis; ds.
OS Escherichia coli DSM498.
XX
FH Key Location/Qualifiers
FT promoter 1..96
FT /tag= a
FT /function= "promoter plac"
FT /evidence= EXPERIMENTAL
FT -35_signal 23..28
FT /tag= b
FT /standard_name= "promoter plac"
FT -10_signal 45..50
FT /tag= c
FT /evidence= EXPERIMENTAL
FT /standard_name= "promoter plac"
FT 105..109
FT /tag= d
FT /evidence= EXPERIMENTAL
FT RBS /standard_name= "biob RBS no. 9"
FT 117..1157
FT /tag= e
FT /product= "biotin synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "biob"
FT /number= 1
FT 1141..1146
FT /tag= f
FT /standard_name= "biob RBS"
FT 1154..2311
FT /tag= g
FT /EC_number= 2.3.1.47
FT /product= "KAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioc"
FT /number= 2
FT /standard_name= "8-amino-7-oxononanoate synthase"
FT 2284..2288
FT /tag= h
FT /standard_name= "bioc RBS"
FT 2295..3050
FT /tag= i
FT /function= "involved in pimeloyl-CoA synthesis"
FT /product= "protein"
FT /gene= "bioc"
FT /number= 3
FT 3030..3033
FT /tag= j
FT /standard_name= "biob RBS"
FT 3043..3753
FT /tag= k
FT /EC_number= 6.3.3.3
FT /product= "PDB synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "biob15"
FT /number= 4
FT /standard_name= "dethiobiotin synthase"
FT 3712..3750
FT /tag= l
FT /note= "biob15 substitution"
FT 3742..3746
FT /tag= m
FT /standard_name= "biob RBS"
FT 3750..5039
FT /tag= n
FT /EC_number= 2.6.1.62
FT /product= "DAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "biob"
FT /number= 5
FT /standard_name= "S-adenosyl-L-methionine: 8-amino-

```

```

FT RBS 5088..5093 7-oxononanoate aminotransferase"
FT /tag= o
FT /standard_name= "ORFI RBS"
FT CDS 5098..5574
FT /tag= p
FT /function= "unknown, involved in biotin synthesis"
FT /product= "protein"
FT /evidence= EXPERIMENTAL
FT /gene= "ORFI"
FT /number= 6
FT terminator 5583..5644
FT /tag= q
FT /standard_name= "rho-independent transcriptional
FT stem_loop 5583..5605
FT /tag= r
FT W09408023-A.
FT 14-APR-1994.
FT 01-OCT-1993; 93WO-EP02688.
FT 02-OCT-1992; 92CH-0003124.
FT 15-JUL-1993; 93CH-0002134.
FT (LONZ ) LONZA AG.
FT Birch O, Brass J, Fuhrmann M, Shaw N;
FT WPI: 1994-135587/16.
FT P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT Biotechnological biotin prodn. using enterobacterial biotin-gene
FT - providing vitamin H in high yield
FT Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
FT The sequence is derived from plasmid pB030A-15/9 contg. the
FT biob, bioc, biob, biob and biob genes responsible for biosynthesis
FT of biotin, arranged in a transcription unit. Microorganisms
FT contg. these DNA fragments or plasmids may be used in the prodn.
FT of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
FT loss of appetite and tiredness.
FT Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other:
XX
XX Query Match 100.0%; Score 25; DB 15; Length 5872;
XX Best Local Similarity 100.0%; Pred. NO. 0.049;
XX Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ggatcgtcgtttgaagcgcagcag 25
DB 179 ggatcgtcgtttgaagcgcagcag 203
RESULT 5
AAA92065/C
ID AAA92065 standard; cDNA; 2515 BP.
XX
AC AAA92065;
XX
XX 04-JAN-2001 (first entry)
XX
XX Human Lhx3c nucleic acid sequence SEQ ID NO:25.
XX
XX Lhx3; LIM-3; P-LIM; identification; characterisation; diagnosis;
XX chromosome 9; pituitary disease; subtelomeric region; mutation;
XX pituitary trophic hormone gene promoter; ss.
XX
XX Homo sapiens.
OS

```

```

XX PN WO200050868-A2.
XX PD 31-AUG-2000.
XX PF 22-FEB-2000; 2000WO-US04424.
XX PR 22-FEB-1999; 99US-0121110.
XX PA (ADRE-) ADVANCED RES & TECHNOLOGY INST.
XX PI Rhodes SJ, Bridwell JL, Meier BC, Parker GE, Price JR;
XX PI Showalter AD, Sloop KW;
XX DR WPI: 2000-594085/56.
XX PT New isolated nucleic acid encoding mammalian Lhx3 for identifying a
XX PT human with a disease, disorder, or condition caused by an altered level
XX PT of expression or binding of Lhx3 -
XX PS Example 4; Fig 19; 239pp; English.
XX CC The present invention describes an isolated nucleic acid (1) encoding a
XX CC mammalian Lhx3. (1) is used in assays to: (1) detect and quantify the
XX CC presence and level of expression of Lhx3, Lhx3a or Lhx3b, in a sample;
XX CC (2) identify a compound that affects expression, the level of expression,
XX CC or the activity of Lhx3, Lhx3a, or Lhx3b in a cell; (3) identify a
XX CC compound that affects binding or Lhx3 to nucleic acid or Lhx3 induction
XX CC of a pituitary trophic hormone gene promoter; (4) identify a human
XX CC afflicted with a disease, disorder, or condition caused by altered
XX CC expression of Lhx3 or altered level of binding of Lhx3 to a nucleic acid;
XX CC and (5) detect a mutation in a Lhx3 allele in a human. The coding region
XX CC of human Lhx3 has been genomically mapped to the subtelomeric region of
XX CC chromosome 9. Lhx3 is also known as P-LIM or LIM-3. The present sequence
XX CC represents the human Lhx3 nucleic acid sequence which is used in an
XX CC example from the present invention.
XX SQ Sequence 2515 BP; 454 A; 819 C; 760 G; 482 T; 0 other;

Query Match 74.4%; Score 18.6; DB 21; Length 2515;
Best Local Similarity 84.0%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggaatcgtcgttgaagcgcagcag 25
DB 183 GGATCTGCTCTCTGAGCCCGCAG 159

RESULT 6
AAA92058/c
ID AAA92058 standard; DNA: 8867 BP.
XX AC AAA92058;
XX DT 04-JAN-2001 (first entry)
XX DE Human Lhx3 genomic DNA SEQ ID NO:22.
XX KW Lhx3; LIM-3; P-LIM; identification; characterisation; diagnosis;
XX KW chromosome 9; pituitary disease; subtelomeric region; mutation;
XX KW pituitary trophic hormone gene promoter; ds.
XX OS Homo sapiens.
XX PN WO200050868-A2.
XX PD 31-AUG-2000.
XX PF 22-FEB-2000; 2000WO-US04424.
XX PR 22-FEB-1999; 99US-0121110.
XX

```

```

PA (ADRE-) ADVANCED RES & TECHNOLOGY INST.
XX PI Rhodes SJ, Bridwell JL, Meier BC, Parker GE, Price JR;
XX PI Showalter AD, Sloop KW;
XX DR WPI: 2000-594085/56.
XX PT New isolated nucleic acid encoding mammalian Lhx3 for identifying a
XX PT human with a disease, disorder, or condition caused by an altered level
XX PT of expression or binding of Lhx3 -
XX PS Claim 30; Fig 26; 239pp; English.
XX CC The present invention describes an isolated nucleic acid (1) encoding a
XX CC mammalian Lhx3. (1) is used in assays to: (1) detect and quantify the
XX CC presence and level of expression of Lhx3, Lhx3a or Lhx3b, in a sample;
XX CC (2) identify a compound that affects expression, the level of expression,
XX CC or the activity of Lhx3, Lhx3a, or Lhx3b in a cell; (3) identify a
XX CC compound that affects binding or Lhx3 to nucleic acid or Lhx3 induction
XX CC of a pituitary trophic hormone gene promoter; (4) identify a human
XX CC afflicted with a disease, disorder, or condition caused by altered
XX CC expression of Lhx3 or altered level of binding of Lhx3 to a nucleic acid;
XX CC and (5) detect a mutation in a Lhx3 allele in a human. The coding region
XX CC of human Lhx3 has been genomically mapped to the subtelomeric region of
XX CC chromosome 9. Lhx3 is also known as P-LIM or LIM-3. The present sequence
XX CC represents the specifically claimed human Lhx3 genomic DNA sequence
XX CC given in the present invention.
XX SQ Sequence 8867 BP; 1462 A; 2874 C; 2940 G; 1589 T; 2 other;

Query Match 74.4%; Score 18.6; DB 21; Length 8867;
Best Local Similarity 84.0%; Pred. No. 36;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggaatcgtcgttgaagcgcagcag 25
DB 4007 GGATCTGCTCTCTGAGCCCGCAG 3983

RESULT 7
AAAX20711
ID AAAX20711 standard; DNA: 808 BP.
XX AC AAAX20711;
XX DT 05-MAY-1999 (first entry)
XX DE Polynucleotide sequence from the genome of Treponema pallidum.
XX KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
XX KW enzyme production; ds.
XX OS Treponema pallidum.
XX PN WO9859034-A2.
XX PD 30-DEC-1998.
XX PF 23-JUN-1998; 98WO-US13041.
XX PR 24-JUN-1997; 97US-0050667.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Fraser CM;
XX PI WPI: 1999-081273/07.
XX DR New isolated Treponema pallidum nucleic acids - used to develop
XX PT products for the detection, diagnosis, characterisation, prevention
XX PT and therapy of T. pallidum infections, particularly syphilis
XX

```

PS Claim 1; Page 864; 1150pp; English.
 CC AAX20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection,
 CC diagnosis, characterisation, prevention and therapy for T. pallidum
 CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to Borrelia infections in animals, and for the
 CC production of biosynthetic products such as enzymes.
 SQ Sequence 808 BP; 204 A; 200 C; 221 G; 180 T; 3 other;
 QY 2 gatctgctgttgaagcgacag 25
 DB 499 gatgtgcttttaagcgacag 522
 Query Match 70.4%; Score 17.6; DB 20; Length 808;
 Best Local Similarity 83.3%; Pred. No. 75;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 RESULT 8
 ID AAC65414/c
 AA085414 standard; DNA; 3561 BP.
 AC AA085414;
 DT 06-OCT-1995 (first entry)
 DE Human TYK2 kinase coding sequence.
 KW JAK family; protein tyrosine kinase; cytokine receptor;
 KW phosphorylation; signal transduction; activation; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 1..3561
 FT /*tag= a
 FT /product= human_TYK2
 FT /note= "nucleotides 817..1680 shown as Fig.3C
 and nucleotides 1681..2544 shown as Fig.3B"
 XX
 PN W09503701-A.
 PD 09-FEB-1995.
 XX
 PF 29-JUL-1994; 94WO-US08676.
 XX
 PR 29-JUL-1993; 93US-0097997.
 XX
 PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 XX
 PI Ihle JN, Quelle FW, Silvenoinen O, Witthuhn BA;
 DR WPI: 1995-081950/11.
 DR P-PSDB: AAB70832.
 XX
 PT Inhibiting a cellular response to a cytokine by inhibiting Jak
 PT kinase - to treat diseases caused by excessive response to
 PT cytokine, e.g. erythrocytosis and other cellular proliferative
 PT diseases
 XX
 PS Claim 29; Fig 3A,C,B,D,E; 167pp; English.
 CC Inhibiting the activity of a jak kinase (pref. Jak1, Jak2, Jak3 or
 CC Tyk2) in a eukaryotic cell is claimed as a method of inhibiting the
 CC biological response of that cell to a cytokine (not IL-3 or
 CC erythropoietin). The present sequence (human TYK2 kinase) was
 CC published by Firmbach-Kraft et al., Oncogene 5:1329-1336 (1990).
 CC It encodes a 1187 amino acid protein which includes the epitopic
 CC sequence SPSEKHFYQKHRLPEPS (amino acids 819-837). Antibodies which
 CC selectively bind this epitope are able to bind Tyk2 without

CC interfering with the activity of the kinase. Such antibodies are
 CC claimed and are useful for detecting and extracting Tyk2.
 SQ Sequence 3561 BP; 699 A; 1112 C; 1098 G; 652 T; 0 other;
 QY 2 gatctgctgttgaagcgacag 25
 DB 3021 GATCTGCTGCGGCGAAGACGACGAG 2998
 Query Match 70.4%; Score 17.6; DB 16; Length 3561;
 Best Local Similarity 83.3%; Pred. No. 94;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 RESULT 9
 ID AAC66246/c
 AAC66246 standard; DNA; 3561 BP.
 AC AAC66246;
 DT 19-FEB-2001 (first entry)
 DE Human tyk2 kinase nucleotide sequence.
 XX
 KW Jak3; kinase; cytokine; cellular response; inhibition; tyk2; ds;
 KW cell proliferation; erythrocytosis.
 OS Homo sapiens.
 XX
 PN US6136595-A.
 PD 24-OCT-2000.
 XX
 PF 18-JUN-1996; 96US-066574.
 XX
 PR 29-JUL-1994; 94US-0282012.
 PR 29-JUL-1993; 93US-0097997.
 PR 09-SEP-1993; 93US-0118968.
 XX
 PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 XX
 PI Silvenoinen O, Witthuhn BA, Ihle J;
 DR WPI: 2000-686080/67.
 DR P-PSDB: AAB35726.
 XX
 PT New DNA encoding Jak3 kinase is useful as cytokine regulator for
 PT treating cell proliferation
 XX
 PS Disclosure; Fig 3; 100pp; English.
 CC This invention relates to DNA encoding a murine Jak3 protein. The amino
 CC acid sequence of the Jak3 protein is given in AAB35715. The Jak family
 CC of kinases are involved in the cellular response to the binding of
 CC cytokines to their respective receptors. Jak3 kinase mediated activation
 CC of some cytokines through their phosphorylation in response to
 CC cytokine-receptor binding. Inhibiting the activity of Jak3 kinase (at the
 CC nucleic acid level with antisense sequences or ribozymes, or at the
 CC protein level with antibodies, kinase inhibitors etc.) is used to treat,
 CC or diagnose, diseases caused by excessive secretion of certain cytokines,
 CC e.g. excessive cell proliferation such as erythrocytosis. Alternatively,
 CC the Jak3 protein and polynucleotide can be used to treat conditions
 CC associated with defective Jak3 activity. The DNA sequence can be used to
 CC produce recombinant Jak3 and this used to raise antibodies useful as
 CC specific inhibitors or to detect or isolate Jak3 without interfering
 CC with its enzymatic activity. The present sequence represents DNA encoding
 CC a tyk2 protein used in the isolation and characterisation of the jak3
 CC protein of the invention.
 SQ Sequence 3561 BP; 699 A; 1113 C; 1097 G; 652 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 3561;
 Best Local Similarity 83.3%; Pred. No. 94;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 gatctgctgttgaagcgacag 25
 |||||
 Db 3021 GATCTGCTGGCGAAGACGACGACG 2998

RESULT 10
 AAQ40480
 ID AAQ40480 standard; cDNA; 15894 BP.
 XX
 AC AAQ40480;
 XX
 DT 19-AUG-1993 (first entry)
 XX
 DE Attenuated measles virus strain AIK-C.
 XX
 KW paramyxoviridae; RNA virus; attenuation; ss.
 XX
 OS Human measles virus.
 XX
 FH Key Location/Qualifiers
 FT CDS 108..1685
 FT /*tag= a
 FT /*label= N
 FT 1807..3330
 FT /*tag= b
 FT /*label= P/C
 FT 3438..4445
 FT /*tag= c
 FT /*label= M
 FT 5458..7110
 FT /*tag= d
 FT /*label= F
 FT 7271..9124
 FT /*tag= e
 FT /*label= H
 FT CDS 9233..15785
 FT /*tag= f
 FT /*label= L
 FT
 FT
 PN EP540135-A.
 XX
 PD 05-MAY-1993.
 XX
 PF 10-MAR-1992; 92EP-0302004.
 XX
 PR 14-OCT-1991; 91JP-0293625.
 XX
 PA (KITA) KITASATO INST.
 PA (KITA) KITASATO KENKYUSHO SH.
 XX
 PI Makino S, Mori T, Sasaki K;
 XX
 DR WPI: 1993-145503/18.
 DR P-PSDB: AAR34537, AAR34538, AAR34539, AAR34540, AAR34541, AAR39592.
 XX
 PT New attenuated measles vaccine virus strain - retains high
 PT immunogenicity with reduced pyrogenicity and having no
 PT neurological complications
 XX
 PS Claim 1; Page 4-37; 47pp; English.
 XX
 CC This sequence represents the entire genome of the attenuated measles
 CC virus strain AIK-C. The virus retains high immunogenicity with
 CC reduced pyrogenicity and no neurological complications.
 XX
 SQ Sequence 15894 BP; 4649 A; 3795 C; 3726 G; 3724 T; 0 other;

Query Match 70.4%; Score 17.6; DB 14; Length 15894;

Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 gatctgctgttgaagcgacag 25
 |||||
 Db 12940 gatctgctgttagaataagcaacag 12963

RESULT 11
 AAV18264
 ID AAV18264 standard; DNA; 15894 BP.
 XX
 AC AAV18264;
 XX
 DT 13-OCT-1998 (first entry)
 XX
 DE Measles virus Edmonston isolate wild-type L protein genomic DNA.
 XX
 KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
 KW single stranded RNA virus; Mononegavirales; ss.
 XX
 OS Measles virus.
 XX
 FH Key Location/Qualifiers
 FT CDS 9234..15785
 FT /*tag= a
 FT /*product= "L protein"
 FT
 PN W09813501-A2.
 XX
 PD 02-APR-1998.
 XX
 PF 19-SEP-1997; 97WO-US16718.
 XX
 PR 27-SEP-1996; 96US-0026823.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
 XX
 DR WPI: 1998-230710/20.
 DR P-PSDB: AAW48703.
 XX
 PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
 PT single stranded RNA virus of order Mononegavirales - having
 PT attenuating mutation in 3' genomic promoter region and RNA
 PT polymerase gene, useful as vaccine to immunise against such virus
 XX
 PS Disclosure; Page 87-97; 426pp; English.
 XX
 CC This sequence encodes the wild-type L protein from Measles virus isolate
 CC Edmonston. This sequence is used in a method which involves the isolation
 CC of recombinantly-generated, attenuated, non-segmented, negative-sense,
 CC single stranded RNA virus of the order Mononegavirales which have at
 CC least 1 attenuating mutation in the 3' genomic promoter region and at
 CC least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
 CC can be used as a vaccine to immunise an individual against such a virus.
 XX
 SQ Sequence 15894 BP; 4639 A; 3800 C; 3737 G; 3718 T; 0 other;

Query Match 70.4%; Score 17.6; DB 19; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 gatctgctgttgaagcgacag 25
 |||||
 Db 12940 gatctgctgttagaataagcaacag 12963

RESULT 12
 AAV18268

```

ID  AAV18268 standard; DNA: 15894 BP.
XX
XX  AAV18268;
AC
XX  13-OCT-1998 (first entry)
DT
XX  Measles virus Rubeovax vaccine L protein genomic DNA.
DE
XX  L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KM  single stranded RNA virus; Mononegavirales; ss.
XX
XX  Measles virus.
OS
XX  Key Location/Qualifiers
FH  9234..15785
FT  /tag= a
FT  /product= "L protein"
XX
XX  W09813501-A2.
PN
XX  02-APR-1998.
PD
XX  19-SEP-1997; 97WO-US16718.
PF
XX  27-SEP-1996; 96US-0026823.
PR
XX  (AMCY ) AMERICAN CYANAMID CO.
PA  (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX  Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
PI  WPI; 1998-230710/20.
DR  P-PSDB; AAW48707.
XX
XX  Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT  single stranded RNA virus of order Mononegavirales - having
PT  attenuating mutation in 3' genomic promoter region and RNA
PT  polymerase gene, useful as vaccine to immunise against such virus
XX
XX  Claim 31; Page 161-172; 426pp; English.
XX
XX  This sequence encodes the L protein from Measles virus Rubeovax
CC  vaccine. This sequence is used in a method which involves the
CC  isolation of recombinantly-generated, attenuated, non-segmented,
CC  negative-sense, single stranded RNA virus of the order Mononegavirales
CC  which have at least 1 attenuating mutation in the 3' genomic promoter
CC  region and at least 1 attenuating mutation in the RNA polymerase gene.
CC  This RNA virus can be used as a vaccine to immunise an individual against
CC  such a virus.
XX
XX  Sequence 15894 BP; 4634 A; 3804 C; 3737 G; 3719 T; 0 other;
SQ

```

```

KM  single stranded RNA virus; Mononegavirales; ss.
XX
XX  Measles virus.
OS
XX  Key Location/Qualifiers
FH  9234..15785
FT  /tag= a
FT  /product= "L protein"
XX
XX  W09813501-A2.
PN
XX  02-APR-1998.
PD
XX  19-SEP-1997; 97WO-US16718.
PF
XX  27-SEP-1996; 96US-0026823.
PR
XX  (AMCY ) AMERICAN CYANAMID CO.
PA  (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX  Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
PI  WPI; 1998-230710/20.
DR  P-PSDB; AAW48708.
XX
XX  Recombinantly-generated, attenuated, non-segmented, negative-sense,
PT  single stranded RNA virus of order Mononegavirales - having
PT  attenuating mutation in 3' genomic promoter region and RNA
PT  polymerase gene, useful as vaccine to immunise against such virus
XX
XX  Claim 31; Page 180-190; 426pp; English.
XX
XX  This sequence encodes the L protein from Measles virus Moraten
CC  vaccine. This sequence is used in a method which involves the
CC  isolation of recombinantly-generated, attenuated, non-segmented,
CC  negative-sense, single stranded RNA virus of the order Mononegavirales
CC  which have at least 1 attenuating mutation in the 3' genomic promoter
CC  region and at least 1 attenuating mutation in the RNA polymerase gene.
CC  This RNA virus can be used as a vaccine to immunise an individual against
CC  such a virus.
XX
XX  Sequence 15894 BP; 4640 A; 3798 C; 3736 G; 3720 T; 0 other;
SQ

```

```

Query Match          70.4%; Score 17.6; DB 19; Length 15894;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Query Match          70.4%; Score 17.6; DB 19; Length 15894;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

OY  2 gatctgctgttgaagcgacag 25
    ||| ||| ||| ||| ||| |||
DB  12940 gatctgctgttagatagcaacag 12963

RESULT 13
AAV18269
ID  AAV18269 standard; DNA: 15894 BP.
XX
XX  AAV18269;
AC
XX  13-OCT-1998 (first entry)
DT
XX  Measles virus Moraten vaccine L protein genomic DNA.
DE
XX  L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KM

```

```

AAV18270
ID  AAV18270 standard; DNA: 15894 BP.
XX
XX  AAV18270;
AC
XX  13-OCT-1998 (first entry)
DT
XX  Measles virus Zagreb vaccine L protein genomic DNA.
DE
XX  L protein; attenuation; non-segmented; negative sense; vaccine; immunity;
KM  single stranded RNA virus; Mononegavirales; ss.
XX
XX  Measles virus.
OS
XX  Key Location/Qualifiers
FH  9234..15785
FT  /tag= a
FT  /product= "L protein"
XX

```

PN WO9813501-A2.
 XX
 PD 02-APR-1998.
 XX
 PF 19-SEP-1997; 97WO-US16718.
 XX
 PR 27-SEP-1996; 96US-0026823.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
 XX
 DR WPI: 1998-230710/20.
 DR P-PSDB; AAM48709.
 XX
 PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
 PT single stranded RNA virus of order Mononegavirales - having
 PT attenuating mutation in 3' genomic promoter region and RNA
 PT polymerase gene, useful as vaccine to immunise against such virus
 XX
 PS Claim 31; Page 199-209; 426pp; English.
 CC
 CC This sequence encodes the L protein from Measles virus Zagreb vaccine.
 CC This sequence is used in a method which involves the isolation of
 CC recombinantly-generated, attenuated, non-segmented, negative-sense,
 CC single stranded RNA virus of the order Mononegavirales which have at
 CC least 1 attenuating mutation in the 3' genomic promoter region and at
 CC least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
 CC can be used as a vaccine to immunise an individual against such a virus.
 CC
 SQ Sequence 15894 BP; 4638 A; 3798 C; 3735 G; 3723 T; 0 other;
 XX
 Query Match 70.4%; Score 17.6; DB 19; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 2 gatctgctgttgaagcgacag 25
 ||||||||| ||| ||| |||
 Db 12940 gatctgctgttagatagaacag 12963
 XX
 RESULT 15
 AAV18271
 ID AAV18271 standard; DNA: 15894 BP.
 XX
 AC AAV18271;
 XX
 DT 13-OCT-1998 (first entry)
 XX
 DE Measles virus Aik-C vaccine L protein genomic DNA.
 XX
 KW L protein; attenuation: non-segmented; negative sense; vaccine; immunity;
 KW single stranded RNA virus; Mononegavirales; ss.
 XX
 OS Measles virus.
 XX
 FH Key Location/Qualifiers
 FH CDS 9234..15785
 FT /*tag= a
 FT /product= "L protein"
 XX
 PN WO9813501-A2.
 XX
 PD 02-APR-1998.
 XX
 PF 19-SEP-1997; 97WO-US16718.
 XX
 PR 27-SEP-1996; 96US-0026823.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
 XX
 DR WPI: 1998-230710/20.
 DR P-PSDB; AAM48709.
 XX
 PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
 PT single stranded RNA virus of order Mononegavirales - having
 PT attenuating mutation in 3' genomic promoter region and RNA
 PT polymerase gene, useful as vaccine to immunise against such virus
 XX
 PS Disclosure; Page 217-228; 426pp; English.
 CC
 CC This sequence encodes the L protein from Measles virus Aik-C vaccine.
 CC This sequence is used in a method which involves the isolation of
 CC recombinantly-generated, attenuated, non-segmented, negative-sense,
 CC single stranded RNA virus of the order Mononegavirales which have at
 CC least 1 attenuating mutation in the 3' genomic promoter region and at
 CC least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
 CC can be used as a vaccine to immunise an individual against such a virus.
 CC
 SQ Sequence 15894 BP; 4644 A; 3793 C; 3731 G; 3726 T; 0 other;
 XX
 Query Match 70.4%; Score 17.6; DB 19; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 2 gatctgctgttgaagcgacag 25
 ||||||||| ||| ||| |||
 Db 12940 gatctgctgttagatagaacag 12963

XX
 PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
 XX
 DR WPI: 1998-230710/20.
 DR P-PSDB; AAM48710.
 XX
 PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
 PT single stranded RNA virus of order Mononegavirales - having
 PT attenuating mutation in 3' genomic promoter region and RNA
 PT polymerase gene, useful as vaccine to immunise against such virus
 XX
 PS Disclosure; Page 217-228; 426pp; English.
 CC
 CC This sequence encodes the L protein from Measles virus Aik-C vaccine.
 CC This sequence is used in a method which involves the isolation of
 CC recombinantly-generated, attenuated, non-segmented, negative-sense,
 CC single stranded RNA virus of the order Mononegavirales which have at
 CC least 1 attenuating mutation in the 3' genomic promoter region and at
 CC least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
 CC can be used as a vaccine to immunise an individual against such a virus.
 CC
 SQ Sequence 15894 BP; 4644 A; 3793 C; 3731 G; 3726 T; 0 other;
 XX
 Query Match 70.4%; Score 17.6; DB 19; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 2 gatctgctgttgaagcgacag 25
 ||||||||| ||| ||| |||
 Db 12940 gatctgctgttagatagaacag 12963
 XX
 RESULT 16
 AAV18265
 ID AAV18265 standard; DNA: 15894 BP.
 XX
 AC AAV18265;
 XX
 DT 13-OCT-1998 (first entry)
 XX
 DE Measles virus 1977 isolate wild-type L protein genomic DNA.
 XX
 KW L protein; attenuation: non-segmented; negative sense; vaccine; immunity;
 KW single stranded RNA virus; Mononegavirales; ss.
 XX
 OS Measles virus.
 XX
 FH Key Location/Qualifiers
 FH CDS 9234..15785
 FT /*tag= a
 FT /product= "L protein"
 XX
 PN WO9813501-A2.
 XX
 PD 02-APR-1998.
 XX
 PF 19-SEP-1997; 97WO-US16718.
 XX
 PR 27-SEP-1996; 96US-0026823.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
 XX
 DR WPI: 1998-230710/20.
 DR P-PSDB; AAV18265.
 XX
 PT Recombinantly-generated, attenuated, non-segmented, negative-sense,
 PT single stranded RNA virus of order Mononegavirales - having
 PT attenuating mutation in 3' genomic promoter region and RNA
 PT polymerase gene, useful as vaccine to immunise against such virus

```
XX Claim 31; Page 105-116; 426bp; English.
PS
CC This sequence encodes the L protein from Measles virus isolate 1977.
CC This sequence is used in a method which involves the isolation
CC of recombinantly-generated, attenuated, non-segmented, negative-sense,
CC single stranded RNA virus of the order Mononegavirales which have at
CC least 1 attenuating mutation in the 3' genomic promoter region and at
CC least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
CC can be used as a vaccine to immunise an individual against such a virus.
SQ Sequence 15894 BP; 4669 A; 3780 C; 3716 G; 3729 T; 0 other;

Query Match          70.4%; Score 17.6; DB 19; Length 15894;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgacag 25
   |||||
Db 12940 gatctgctgtgaataagcaacag 12963

RESULT 17
AAV18266
ID AAV18266 standard; DNA; 15894 BP.
XX
AC AAV18266;
XX
DT 13-OCT-1998 (first entry)
DE Measles virus 1983 isolate wild-type L protein genomic DNA.
XX
XX L protein; attenuated; non-segmented; negative sense; vaccine; immunity;
KM single stranded RNA virus; Mononegavirales; ss.
XX
XX Measles virus.
XX
XX Key Location/Qualifiers
FT CDS 9234..15785
FT /*tag= a
FT /product= "L protein"
XX
XX W09813501-A2.
XX
XX 02-APR-1998.
XX
XX 19-SEP-1997; 97MO-US16718.
XX
XX 27-SEP-1996; 96US-0026823.
XX
XX (AMCY ) AMERICAN CYANAMID CO.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;
XX WPI: 1998-230710/20.
XX P-PSDB: AAM48705.
XX
XX Recombinantly-generated, attenuated, non-segmented, negative-sense,
XX single stranded RNA virus of order Mononegavirales - having
XX attenuating mutation in 3' genomic promoter region and RNA
XX polymerase gene, useful as vaccine to immunise against such virus
XX
XX Claim 31; Page 124-134; 426bp; English.
XX
XX This sequence encodes the wild-type L protein from Measles virus isolate
XX 1983. This sequence is used in a method which involves the isolation
XX of recombinantly-generated, attenuated, non-segmented, negative-sense,
XX single stranded RNA virus of the order Mononegavirales which have at
XX least 1 attenuating mutation in the 3' genomic promoter region and at
XX least 1 attenuating mutation in the RNA polymerase gene. This RNA virus
XX can be used as a vaccine to immunise an individual against such a virus.
```

```
XX Sequence 15894 BP; 4639 A; 3791 C; 3729 G; 3735 T; 0 other;
SQ

Query Match          70.4%; Score 17.6; DB 19; Length 15894;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgacag 25
   |||||
Db 12940 gatctgctgtgaataagcaacag 12963

RESULT 18
AAZ22901
ID AAZ22901 standard; DNA; 15894 BP.
XX
AC AAZ22901;
XX
DT 10-JAN-2000 (first entry)
DE Nucleotide sequence of wild type measles virus Edmonston strain.
XX
XX Measles virus; attenuated; human respiratory syncytial virus; RSV;
XX mutation; vaccine; immunization; measles; RSV subgroup B; ss.
XX
XX Measles virus.
XX
XX W09949017-A2.
XX
XX 30-SEP-1999.
XX
XX 22-MAR-1999; 99WO-US06225.
XX
XX 26-MAR-1998; 98US-0079466.
XX
XX (AMCY ) AMERICAN CYANAMID CO.
XX
XX Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;
XX WPI: 1999-580441/49.
XX
XX New vaccines for measles and respiratory syncytial virus (RSV)
XX
XX Disclosure; Page 86-92; 171pp; English.
XX
XX The invention provides isolated, recombinantly-generated, attenuated
XX measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
XX (II). The attenuated measles virus has at least 1 of the following
XX attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
XX or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
XX Leu439Pro; or (3) in the C gene at residues Ala73Val, Met104Thr, or
XX Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
XX The attenuated RSV has an attenuating mutation in the M gene-end signal
XX comprising Thr4199Cys. (I) is useful as a vaccine for immunizing against
XX measles. (II) is useful as a vaccine for immunizing and giving protection
XX against RSV subgroup B. Compositions comprising transcriptional vector
XX comprising an isolated nucleic acid molecule encoding a genome or
XX antigenome of (I) or (II), are useful for producing infectious attenuated
XX measles virus or RSV subgroup B virus. Current vaccines for measles and
XX RSV do not provide 100 % protection, and only give short-lived immunity.
XX Other vaccines give unfavorable immune responses or adverse reactions.
XX
XX Sequence 15894 BP; 4639 A; 3800 C; 3737 G; 3718 T; 0 other;
SQ

Query Match          70.4%; Score 17.6; DB 20; Length 15894;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgacag 25
   |||||
Db 12940 gatctgctgtgaataagcaacag 12963
```

RESULT 19
AA222902 standard; DNA; 15894 BP.
XX
AC AA222902;
XX
DT 10-JAN-2000 (first entry)
XX
DE Nucleotide sequence of wild type measles virus 1977 strain.
XX
KM Measles virus; attenuated; human respiratory syncytial virus: RSV;
XX mutation; vaccine; immunization; measles; RSV subgroup B; ss.
XX
OS Measles virus.
XX
PN W09949017-A2.
XX
PD 30-SEP-1999.
XX
PF 22-MAR-1999; 99WO-US06225.
XX
PR 26-MAR-1998; 98US-0079466.
XX
PA (AMCY) AMERICAN CYANAMID CO.
XX
PI Udem SA, Sidhu MS, Randolph VB, Buonaugurio DA;
XX WPI; 1999-580441/49.
XX
PT New vaccines for measles and respiratory syncytial virus (RSV) -
XX
PS Disclosure; Page 92-97; 171pp; English.
XX
CC The invention provides isolated, recombinantly-generated, attenuated
CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
CC (II). The attenuated measles virus has at least 1 of the following
CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr or
CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
CC The attenuated RSV has an attenuating mutation in the M gene-end signal
CC comprising Thr4199Cys. (I) is useful as a vaccine for immunizing against
CC measles. (II) is useful as a vaccine for immunizing against
CC against RSV subgroup B. Compositions comprising transcriptional vector
CC comprising an isolated nucleic acid molecule encoding a genome or
CC antigenome of (I) or (II), are useful for producing infectious attenuated
CC measles virus or RSV subgroup B virus. Current vaccines for measles and
CC RSV do not provide 100 % protection, and only give short-lived immunity.
CC Other vaccines give unfavorable immune responses or adverse reactions.
XX
SQ Sequence 15894 BP; 4669 A; 3780 C; 3716 G; 3729 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgacag 25
DB 12940 gatctgctgttagaataagacag 12963
|||||

RESULT 20
AA222903 standard; DNA; 15894 BP.
XX
AC AA222903;
XX
DT 10-JAN-2000 (first entry)
XX
DE Nucleotide sequence of wild type measles virus 1983 strain.

XX
KM Measles virus; attenuated; human respiratory syncytial virus; RSV;
XX mutation; vaccine; immunization; measles; RSV subgroup B; ss.
XX
OS Measles virus.
XX
PN W09949017-A2.
XX
PD 30-SEP-1999.
XX
PF 22-MAR-1999; 99WO-US06225.
XX
PR 26-MAR-1998; 98US-0079466.
XX
PA (AMCY) AMERICAN CYANAMID CO.
XX
PI Udem SA, Sidhu MS, Randolph VB, Buonaugurio DA;
XX WPI; 1999-580441/49.
XX
PT New vaccines for measles and respiratory syncytial virus (RSV) -
XX
PS Disclosure; Page 97-103; 171pp; English.
XX
CC The invention provides isolated, recombinantly-generated, attenuated
CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
CC (II). The attenuated measles virus has at least 1 of the following
CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
CC Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr or
CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
CC The attenuated RSV has an attenuating mutation in the M gene-end signal
CC comprising Thr4199Cys. (I) is useful as a vaccine for immunizing against
CC measles. (II) is useful as a vaccine for immunizing against
CC against RSV subgroup B. Compositions comprising transcriptional vector
CC comprising an isolated nucleic acid molecule encoding a genome or
CC antigenome of (I) or (II), are useful for producing infectious attenuated
CC measles virus or RSV subgroup B virus. Current vaccines for measles and
CC RSV do not provide 100 % protection, and only give short-lived immunity.
CC Other vaccines give unfavorable immune responses or adverse reactions.
XX
SQ Sequence 15894 BP; 4639 A; 3791 C; 3729 G; 3735 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgacag 25
DB 12940 gatctgctgttagaataagacag 12963
|||||

RESULT 21
AA222905 standard; DNA; 15894 BP.
XX
AC AA222905;
XX
DT 10-JAN-2000 (first entry)
XX
DE Nucleotide sequence of RubioVax vaccine measles virus genome.
XX
KM Measles virus; attenuated; human respiratory syncytial virus; RSV;
XX mutation; vaccine; immunization; measles; RSV subgroup B; ss.
XX
OS Synthetic.
XX
PN W09949017-A2.
XX
PD 30-SEP-1999.

PF 22-MAR-1999: 99WO-US06225.
 XX
 PR 26-MAR-1998: 98US-0079466.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI udem SA, Sidhu MS, Randolph VB, Buonaugurio DA;
 XX
 DR WPI: 1999-580441/49.
 XX
 PR New vaccines for measles and respiratory syncytial virus (RSV)
 PT
 PS Disclosure: Page 109-114; 171pp; English.
 XX
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene at residues Ala73Val, Met104Thr or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal
 CC comprising Thr4199Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 XX
 SQ Sequence 15894 BP: 4634 A; 3804 C; 3737 G; 3719 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgacag 25
 |||||
 DB 12940 gatctgctgtgaatagacacag 12963

RESULT 22
 AA222907 standard; DNA: 15894 BP.
 ID AA222907;
 AC AA222907;
 XX
 DT 10-JAN-2000 (first entry)
 DE Nucleotide sequence of Zagreb vaccine measles virus genome.
 XX
 KW Measles virus; attenuated; human respiratory syncytial virus; RSV;
 KM mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 XX
 OS Synthetic.
 OS Measles virus.
 XX
 PN WO949017-A2.
 XX
 PD 30-SEP-1999.
 XX
 PF 22-MAR-1999: 99WO-US06225.
 XX
 PR 26-MAR-1998: 98US-0079466.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI udem SA, Sidhu MS, Randolph VB, Buonaugurio DA;
 XX
 DR WPI: 1999-580441/49.
 XX

PT New vaccines for measles and respiratory syncytial virus (RSV) -
 XX
 XX Disclosure: Page 120-126; 171pp; English.
 PS
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene at residues Ala73Val, Met104Thr, or
 CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
 CC The attenuated RSV has an attenuating mutation in the M gene-end signal
 CC comprising Thr4199Cys. (I) is useful as a vaccine for immunizing against
 CC measles. (II) is useful as a vaccine for immunizing and giving protection
 CC against RSV subgroup B. Compositions comprising transcriptional vector
 CC comprising an isolated nucleic acid molecule encoding a genome or
 CC antigenome of (I) or (II), are useful for producing infectious attenuated
 CC measles virus or RSV subgroup B virus. Current vaccines for measles and
 CC RSV do not provide 100 % protection, and only give short-lived immunity.
 CC Other vaccines give unfavorable immune responses or adverse reactions.
 XX
 SQ Sequence 15894 BP: 4638 A; 3798 C; 3735 G; 3723 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgacag 25
 |||||
 DB 12940 gatctgctgtgaatagacacag 12963

RESULT 23
 AA222908 standard; DNA: 15894 BP.
 ID AA222908;
 AC AA222908;
 XX
 DT 10-JAN-2000 (first entry)
 DE Nucleotide sequence of Aik-C vaccine measles virus genome.
 XX
 KW Measles virus; attenuated; human respiratory syncytial virus; RSV;
 KM mutation; vaccine; immunization; measles; RSV subgroup B; ss.
 XX
 OS Synthetic.
 OS Measles virus.
 XX
 PN WO949017-A2.
 XX
 PD 30-SEP-1999.
 XX
 PF 22-MAR-1999: 99WO-US06225.
 XX
 PR 26-MAR-1998: 98US-0079466.
 XX
 PA (AMCY) AMERICAN CYANAMID CO.
 XX
 PI udem SA, Sidhu MS, Randolph VB, Buonaugurio DA;
 XX
 DR WPI: 1999-580441/49.
 XX
 PR New vaccines for measles and respiratory syncytial virus (RSV) -
 PT
 PS Disclosure: Page 126-131; 171pp; English.
 XX
 CC The invention provides isolated, recombinantly-generated, attenuated
 CC measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
 CC (II). The attenuated measles virus has at least 1 of the following
 CC attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
 CC or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
 CC Leu439Pro; or (3) in the C gene at residues Ala73Val, Met104Thr, or

CC Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
CC The attenuated RSV has an attenuating mutation in the M gene-end signal
CC comprising Thr419Gcys. (I) is useful as a vaccine for immunizing against
CC measles. (II) is useful as a vaccine for immunizing and giving protection
CC against RSV subgroup B. Compositions comprising transcriptional vector
CC comprising an isolated nucleic acid molecule encoding a genome or
CC antigenome of (I) or (II), are useful for producing infectious attenuated
CC measles virus or RSV subgroup B virus. Current vaccines for measles and
CC RSV do not provide 100 % protection, and only give short-lived immunity.
CC Other vaccines give unfavorable immune responses or adverse reactions.
CC
SQ Sequence 15894 BP; 4644 A; 3793 C; 3731 G; 3726 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15894;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgcagcag 25
|||||
DB 12940 gatctgctgttagatagaacacag 12963

RESULT 24
AAZ22906
ID AAZ22906 standard; DNA; 15895 BP.
XX
AC AAZ22906;
XX
DT 10-JAN-2000 (first entry)
XX
DE Nucleotide sequence of Moraten vaccine measles virus genome.
XX
KM Measles virus; attenuated; human respiratory syncytial virus; RSV;
XX mutation; vaccine; immunization; measles; RSV subgroup B; ss.
XX
OS Synthetic.
XX Measles virus.
XX
PN WO949017-A2.
XX
PD 30-SEP-1999.
XX
PE 22-MAR-1999; 99WO-US06225.
XX
PR 26-MAR-1998; 98US-0079466.
XX
PA (AMCY) AMERICAN CYANAMID CO.
XX
PI Udem SA, Sidhu MS, Randolph VB, Buonagurio DA;
XX
DR WPI: 1999-580441/49.
XX
PT
XX
PS New vaccines for measles and respiratory syncytial virus (RSV) -
XX
XX Disclosure; Page 114-120; 171pp; English.

The invention provides isolated, recombinantly-generated, attenuated
measles virus (I) and human respiratory syncytial virus (RSV) subgroup B
(II). The attenuated measles virus has at least 1 of the following
attenuating mutations: (1) in the N gene, at residue Glu129Lys, Glu148Gly
or Ser479Thr; (2) in the P gene, at residues Glu225Gly, Cys275Tyr or
Leu439Pro; or (3) in the C gene, at residues Ala73Val, Met104Thr, or
Ser134Tyr; or (4) at the F gene-end signal, at nucleotide Thr7243Cys.
The attenuated RSV has an attenuating mutation in the M gene-end signal
comprising Thr419Gcys. (I) is useful as a vaccine for immunizing against
measles. (II) is useful as a vaccine for immunizing and giving protection
against RSV subgroup B. Compositions comprising transcriptional vector
comprising an isolated nucleic acid molecule encoding a genome or
antigenome of (I) or (II), are useful for producing infectious attenuated
measles virus or RSV subgroup B virus. Current vaccines for measles and
RSV do not provide 100 % protection, and only give short-lived immunity.
Other vaccines give unfavorable immune responses or adverse reactions.

XX
SQ Sequence 15895 BP; 4641 A; 3798 C; 3736 G; 3720 T; 0 other;

Query Match 70.4%; Score 17.6; DB 20; Length 15895;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgcagcag 25
|||||
DB 12941 gatctgctgttagatagaacacag 12964

RESULT 25
AAA59384
ID AAA59384 standard; DNA; 801 BP.
XX
AC AAA59384;
XX
DT 07-NOV-2000 (first entry)
XX
DE DNA clone encoding a murine guanine nucleotide releasing factor 4.
XX
KM guanine nucleotide releasing factor 4; GRF-4; Ras activator;
XX Nedd4 ubiquitination; cell metabolism; cell proliferation; cancer;
XX cell differentiation; cell transformation; neuronal disorder; ss.
XX
OS Mus sp.
XX
FH Key
FH Location/Qualifiers
FT CDS
FT 1..801
FT /*tag- a
FT /note- "contains 5 internal stop codons and encodes
FT AAB07793; no termination codon given"
FT
FT 2..799
FT /note- b
FT /*tag- b
FT /note- "contains 8 internal stop codons and encodes
FT AAB07794; no termination codon given"
FT CDS
FT 3..800
FT /*tag- c
FT /note- "contains 5 internal stop codons and encodes
FT AAB07795; no termination codon given"
XX
PN WO200043510-A2.
XX
PD 27-JUL-2000.
XX
PE 20-JAN-2000; 2000WO-CA00042.
XX
PR 20-JAN-1999; 99CA-2259830.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Rolin D, Pham N;
XX
DR WPI: 2000-499228/44.
XX
DR P-PSDB; AAB07793, AAB07794, AAB07795.
XX
PT Nucleic acids encoding guanine nucleotide releasing factor-4 useful for
XX the treatment of cancers and neuronal disorders -
XX
XX Disclosure; Fig 19B; 89pp; English.

The present sequence encodes a murine guanine nucleotide releasing
factor (GRF)-4 (Ras activator) polypeptide. GRF4 activates Ras both
in vitro and in vivo. It directly binds cyclic adenosine monophosphate
(cAMP) directly via its cAMP-BD (cAMP/guanine monophosphate (cGMP)
binding domain). GRF4 directly connects cAMP-generating (e.g. G protein
coupled receptors) or cGMP-generating pathways to Ras. GRF4 activates
Ras in response to elevation of intracellular cAMP and/or cGMP. GRF4
is a target for Nedd4 ubiquitination as it binds Nedd4. Activation of
the Ras signalling pathway controls numerous cellular functions, such as
cell metabolism, proliferation, differentiation and transformation.

CC Therefore modulation of Ras activity may provide a mechanism for
CC controlling diseases. GRF4 polynucleotides and polypeptides may be used
CC in the treatment of diseases associated with inappropriate GRF4
CC expression and activity such as cancers and neuronal disorders. The
CC GRF4 polypeptides may be used as antigens in the production of
CC antibodies against GRF4 and in assays to identify modulators (agonists
CC and antagonists) of GRF4 expression and activity. The anti-GRF4
CC antibodies and GRF4 antagonists may also be used to down regulate GRF4
CC expression and activity. Inhibition of Ras can reduce cellulose
CC proliferation and cancers.

XX
SQ Sequence 801 BP; 173 A; 266 C; 232 G; 130 T; 0 other;

Query Match 69.6%; Score 17.4; DB 21; Length 801;
Best Local Similarity 94.7%; Pred. No. 92;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 4 tctgctgttgaagcgagcag 22
|||||
Db 525 tctgctgttgaagcgagcag 543

RESULT 26

AAK16338
ID AAK16338 standard; cDNA; 1970 BP.

XX
AC AAK16338;

XX 26-APR-1999 (first entry)

XX Mouse OPF #16 encoding cDNA.

XX Mouse; OPF #16; differentiation inducing activity; bone marrow cell;
XX hydroxyapatite decomposition; ds.

XX OS

XX Mus sp.

XX Key Location/Qualifiers

XX FT 270..953
XX FT /*tag= a
XX FT /product= "OPF #16"

XX JPI1028091-A.

XX 02-FEB-1999.

XX 10-JUL-1997; 97JP-0202201.

XX 10-JUL-1997; 97JP-0202201.

XX (SUMU) SUMITOMO SEIYAKU KK.

XX WPI: 1999-174353/15.

XX P-PSDB; AAW94508.

XX A new protein OPF u 16 and gene - useful for differentiation of bone
XX marrow cells

XX Claim 2; Page 11-12; 13pp; Japanese.

XX The present sequence encodes an OPF protein isolated from mouse,
CC designated OPF #16. OPF #16 has differentiation inducing activity from
CC a bone marrow cell to a cell having hydroxyapatite decomposing activity.
CC OPF genes and proteins are useful for screening for inhibitors against
CC differentiation from bone marrow cells.

XX Sequence 1970 BP; 471 A; 507 C; 623 G; 369 T; 0 other;

Query Match 68.8%; Score 17.2; DB 20; Length 1970;
Best Local Similarity 86.4%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 tctgctgttgaagcgagcag 25
|||||
Db 1806 tctgctgttgaagcgagcag 1827

RESULT 27

AAK44857/c
ID AAK44857 standard; DNA; 2126 BP.

XX
AC AAK44857;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 44393.

XX Hybridisation assay; genetic mapping; gene expression control;

XX protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130510.

XX 23-APR-1999; 99US-0130891.

XX 28-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132863.

XX 11-MAY-1999; 99US-0134256.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

XX 07-JUN-1999; 99US-0138094.

XX 08-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

XX 14-JUN-1999; 99US-0139119.

XX 16-JUN-1999; 99US-0139452.

XX 16-JUN-1999; 99US-0139453.

```
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144337.
PR 19-JUL-1999; 99US-0144338.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144352.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145115.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150864.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151308.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 29-SEP-1999; 99US-0156458.
PR 04-OCT-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 06-OCT-1999; 99US-0157153.
PR 07-OCT-1999; 99US-0157865.
PR 08-OCT-1999; 99US-0158029.
PR 12-OCT-1999; 99US-0158332.
PR 13-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 68.8%; Score 17.2; DB 21; Length 2126;
Best Local Similarity 86.4%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

AA256969/C
 ID AA256969 standard; cDNA; 2203 BP.
 XX
 AC AA256969;
 XX
 DT 12-MAY-2000 (first entry)
 XX
 DE Arabidopsis ATP citrate lyase (ACL) B-2 subunit encoding cDNA.
 XX
 KW Arabidopsis; plant plastid; acetyl CoA synthetase; E3 subunit; ACS;
 KM plant plastidic pyruvate dehydrogenase; PPDH; ATP citrate lyase; ACL;
 KM pyruvate decarboxylase; PDC; aldehyde dehydrogenase; ALDH; acetyl CoA;
 KM fatty acid; flavonoid; enzyme; phytochemical; pyruvate decarboxylase;
 KM acetyl CoA hydrolase; mitochondrial pyruvate dehydrogenase; ss.
 OS
 XX Arabidopsis sp.
 XX
 PN WO200000619-A2.
 XX
 PD 06-JAN-2000.
 XX
 PF 25-JUN-1999; 99WO-US14382.
 XX
 PR 26-JUN-1998; 98US-0090717.
 XX
 PA (IOWA) UNIV IOWA STATE RES FOUND INC.
 XX
 PI Nikolaou BJ, Wurtele ES, Oliver DJ, Behal R, Schnable PS, Ke J;
 PI Johnson JL, Allred CC, Fatland B, Lutziger I, Wen T;
 PI WPI: 2000-160678/14.
 DR P-PSDB; AAY67408.
 XX
 PT Novel acetyl CoA synthetase (ACS), plastidic pyruvate dehydrogenase
 PT (PPDH), ATP citrate lyase (ACL), pyruvate decarboxylase (PDC) and
 PT aldehyde dehydrogenase (ALDH) polypeptides used to alter acetyl CoA
 PT levels in plants -
 XX
 PS Claim 24; Fig 6; 79pp; English.
 XX
 CC The invention provides nucleic acids encoding Arabidopsis plastidic
 CC acetyl CoA synthetase (ACS), various subunits (specifically the E3
 CC subunit) of plant plastidic pyruvate dehydrogenase (PPDH), the A and B
 CC subunits of a plant ATP citrate lyase (ACL), Arabidopsis pyruvate
 CC decarboxylase (PDC), Arabidopsis aldehyde dehydrogenase (ALDH),
 CC specifically ALDH-2 and ALDH-4. The polypeptides can be expressed by
 CC standard recombinant methodology. The ACS, PPDH, ACL, PDC and ALDH
 CC polypeptides, methods and nucleic acid molecules of the invention are
 CC used to alter the level of acetyl CoA in a plant or plant cell, tissues
 CC or organs. A decrease in acetyl CoA is expected to affect the
 CC biosynthesis of very long chain fatty acids and flavonoids. The enzymes
 CC may also be used for in vitro synthesis of acetyl CoA, which in turn can
 CC be used to produce acetyl CoA phytochemicals (plastidic ACS, PPDH, ACL,
 CC pyruvate decarboxylase, acetyl CoA hydrolase, mitochondrial pyruvate
 CC dehydrogenase and aldehyde dehydrogenase are involved in the biosynthesis
 CC of acetyl CoA.
 CC
 XX
 SQ Sequence 2203 BP; 564 A; 480 C; 494 G; 665 T; 0 other;

Query Match 68.8%; Score 17.2; DB 21; Length 2203;
 Best Local Similarity 86.4%; Pred. No. 1.3e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctgctgttgaagcgagca 24
 ||| ||||| ||||| |||||
 DB 332 ATCCGCTGTGGATCGCAGCA 311

RESULT 29
 AAX20559/C
 ID AAX20559 standard; DNA; 5504 BP.
 XX

AC AAX20559;
 XX
 DT 05-MAY-1999 (first entry)
 XX
 DE Polynucleotide sequence from the genome of Treponema pallidum.
 XX
 KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
 KM enzyme production; ds.
 XX
 OS Treponema pallidum.
 XX
 PN WO9859034-A2.
 XX
 PD 30-DEC-1998.
 XX
 PF 23-JUN-1998; 98WO-US13041.
 XX
 PR 24-JUN-1997; 97US-0050667.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fraser CM;
 PI WPI: 1999-081273/07.
 DR
 XX
 PT New isolated Treponema pallidum nucleic acids - used to develop
 PT products for the detection, diagnosis, characterization, prevention
 PT and therapy of T. pallidum infections, particularly syphilis
 PT
 PS Claim 1; Page 519-522; 1150pp; English.
 XX
 CC AAX20500-21243 represent polynucleotide sequences from the genome of
 CC Treponema pallidum. The sequences can be used for detection,
 CC diagnosis, characterization, prevention and therapy for T. pallidum
 CC infections, particularly syphilis. They can also be used for detecting
 CC diseases related to Borrelia infections in animals, and for the
 CC production of biosynthetic products such as enzymes.
 CC
 XX
 SQ Sequence 5504 BP; 1474 A; 1764 C; 1181 G; 1076 T; 9 other;

Query Match 68.8%; Score 17.2; DB 20; Length 5504;
 Best Local Similarity 86.4%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgagc 23
 ||| ||||| ||||| |||||
 DB 1370 GGTGGGCTGTGAGCGCAGC 1349

RESULT 30
 AAX61129/C
 ID AAX61129 standard; DNA; 252 BP.
 XX
 AC AAX61129;
 XX
 DT 12-OCT-2000 (first entry)
 XX
 DE SEN virus genome fragment SEQ ID NO: 40.
 XX
 KW SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease;
 KM proliferative disorder; hepatopathy; hepatitis; viral infection;
 KM vaccination; gene therapy; ds.
 XX
 OS Hepatitis virus.
 XX
 PN WO200028039-A2.
 XX
 PD 18-MAY-2000.
 XX
 PF 09-NOV-1999; 99WO-EP08566.
 XX
 PR 10-NOV-1998; 98IT-MI02437.

```
PR 30-APR-1999; 99IT-MI00923.
PR 14-MAY-1999; 99EP-0830298.
PR 16-JUL-1999; 99EP-0113932.
XX
PA (DIAS-) DIASORIN SRL.
XX
PI Print D, Fioridalisi G, Mantero GL, Mattioli S, Sottini A;
PI Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;
XX
DR WPI: 2000-376551/32.
XX
PT Nucleic acids representing the genome of the SPN virus (SENV) and
PT encoded proteins, useful for treatment of hepatopathies, inflammatory
PT diseases and proliferative disorders such as cancer.
XX
PS Claim 1; Page 316; 392pp; English.
XX
CC The present invention is concerned with the sequence of the genome of the
CC SEN virus (SENV), and the proteins encoded by it. SENV is thought to be
CC the cause of hepatopathies which are not linked to the presence of the
CC hepatitis A, B and E viruses in man. The genome and proteins of this
CC virus can be used in gene therapy and vaccination against the virus,
CC which also causes disorders of the gastrointestinal tract, including
CC Crohn's disease and lupus erythematosus, inflammatory diseases, and
CC proliferative disorders such as cancer.
XX
SQ Sequence 252 BP; 84 A; 79 C; 48 G; 41 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 252;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgacag 25
    ||| ||||| ||||| ||||| |||
Db 204 GGAGCTGCTGCTGATGAGCAGCTG 180

RESULT 31
AAC75304/c
ID AAC75304 standard; CDNA: 312 BP.
XX
AC AAC75304;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF859 polynucleotide sequence SEQ ID NO:1717.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-0508621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
```

```
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach M;
PI
XX
DR WPI: 2000-602362/57.
DR P-PSDB; AAB41095.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
PS Claim 5; Page 1363; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, diabetes mellitus,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 312 BP; 88 A; 82 C; 86 G; 55 T; 1 other;

Query Match 68.0%; Score 17; DB 21; Length 312;
Best Local Similarity 80.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgacag 25
    ||||| || ||| ||||| |||
Db 256 GGATCTGTGCTGATCCGACAGCAG 232

RESULT 32
AAC07943
ID AAC07943 standard; CDNA: 394 BP.
XX
AC AAC07943;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 12018.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; CDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
```

PA (GEST) GENSFT.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI: 2000-500381/45.
 DR
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 1: SEQ ID 12018; 71pp + CD-ROM; English.
 XX
 CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.
 CC
 SQ Sequence 394 BP; 71 A; 118 C; 115 G; 83 T; 7 other;
 XX
 SQ
 Query Match 68.0%; Score 17; DB 21; Length 394;
 Best Local Similarity 80.0%; Pred. No. 1.3e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 ggatctgctgttgaagcgcagcag 25
 ||| ||||| | ||| ||||| |
 Db 29 ggacctgctgtcgtgaagcgcagcgcg 53

RESULT 33
 AAC93432
 ID AAC93432 standard; cDNA; 1579 BP.
 XX
 AC AAC93432:
 XX
 DT 16-FEB-2001 (first entry)
 XX
 DE Human secreted protein gene 11 SEQ ID NO:21.
 XX
 KW Human; secreted protein; immunosuppressive; antiarthritis; antirheumatic;
 KW antiproliferative; cytoskeletal; cardiact; vasotropic; cerebroprotective;
 KW neurotropic; neuroprotective; antibacterial; viruticidal; fungicide;
 KW ophthalmological; vulnerrary; autoimmune disease; Rheumatoid arthritis;
 KW hyperproliferative disorders; cancer; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
 KW Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200061625-A1.
 XX
 PD 19-OCT-2000.
 XX
 PF 06-APR-2000; 2000MO-US08981.
 XX
 XX 09-APR-1999; 99US-0128701.
 PR 20-JAN-2000; 2000US-0177166.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis G;
 XX

DR WPI: 2000-619226/59.
 DR P-PSDB: AAB51735.
 XX
 PT New nucleic acid molecules encoding 48 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 XX
 PS Claim 1; Page 417; 500pp; English.
 XX
 CC Polynucleotide sequences AAC93422 - AAC93449 represent cDNA encoding
 CC human secreted proteins AAB51724 - AAB51777. Sequences AAB51778 -
 CC AAB51825 represent alternative polypeptides encoded by the genes, and
 CC amino acid sequences to which they are homologous. The genes and proteins
 CC have activities dependent on the tissues and cells in which they are
 CC expressed. Examples of their activities include immunosuppressive;
 CC antiarthritis; antirheumatic; antiproliferative; cytoskeletal; cardiact;
 CC vasotropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
 CC viruticidal; fungicide; ophthalmological; and vulnerrary. The secreted
 CC proteins, polynucleotides, antagonists and agonists may be useful in
 CC treating, preventing and/or diagnosing diseases and disorders such as
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
 CC angioneuosis, nervous system disorders e.g. Alzheimer's disease,
 CC infections caused by bacteria, viruses and fungi and ocular disorders
 CC e.g. corneal infection. The polypeptides can also be used to aid wound
 CC healing and epithelial cell proliferation, to prevent skin aging due to
 CC sunburn, to maintain organs before transplantation, for supporting cell
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
 CC polypeptides can also be used as a food additive or preservative to
 CC increase or decrease storage capabilities, fat content, lipid, protein,
 CC carbohydrate, vitamins, minerals, cofactors and other nutritional
 CC components. Oligonucleotide AAC93413 - AAC93421 and peptide AAB51723 are
 CC used in the isolation and characterisation of the proteins and
 CC polynucleotides of the invention.
 CC
 SQ Sequence 1579 BP; 344 A; 365 C; 373 G; 488 T; 9 other;
 XX
 SQ
 Query Match 68.0%; Score 17; DB 21; Length 1579;
 Best Local Similarity 80.0%; Pred. No. 1.5e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 ggatctgctgttgaagcgcagcag 25
 ||||| | ||| ||||| |
 Db 1186 ggatctgctgtcgtgacgcgcagcag 1210

RESULT 34
 AAC76996
 ID AAC76996 standard; cDNA; 1711 BP.
 XX
 AC AAC76996:
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF2551 polynucleotide sequence SEQ ID NO:5101.
 XX
 KW Human; open reading frame; ORFX; detection; cytoskeletal; hepatotropic;
 KW vulnerrary; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritis; immunosuppressant; cardiact;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX

OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB42787.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 4289-4290; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferative; antiparasitoid; neuroprotective;
 CC osteoplastic; anticonvulsant; antiallergic; immunosuppressive;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antihypertensive; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hyperextension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX
 SO Sequence 1711 BP; 341 A; 518 C; 498 G; 354 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 1711;
 Best Local Similarity 80.0%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggatcgtctgttgaagcagcagc 25
 ||| ||||| ||| |||||
 DB 1 ggactctgtctgaagcagcagc 25

RESULT 35
 ID AAA61127 standard; DNA; 2040 BP.
 AC AAA61127;
 XX
 DT 12-OCT-2000 (first entry)
 XX
 DE SEN virus genome fragment SEQ ID NO: 38.
 XX
 KW SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease;
 KW proliferative disorder; hepatopathy; hepatitis; viral infection;
 KW vaccination; gene therapy; ds. ;

XX
 OS Hepatitis virus.
 XX
 PN WO200028039-A2.
 XX
 PD 18-MAY-2000.
 XX
 PF 09-NOV-1999; 99WO-EP08566.
 XX
 PR 10-NOV-1998; 98IT-MT02437.
 PR 30-APR-1999; 99IT-MT00923.
 PR 14-MAY-1999; 99EP-0830298.
 PR 16-JUL-1999; 99EP-0113932.
 XX
 PA (DIAS-) DIASORIN SRL.
 XX
 PI Primi D, Fioridalisi G, Mantero GL, Mattioli S, Sottini A;
 PI Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;
 XX
 DR WPI: 2000-376551/32.
 XX
 PT Nucleic acids representing the genome of the SEN virus (SENV) and
 PT encoded proteins; useful for treatment of hepatopathies, inflammatory
 PT diseases and proliferative disorders such as cancer -
 XX
 PS Claim 1; Page 315-316; 392pp; English.
 XX
 CC The present invention is concerned with the sequence of the genome of the
 CC SEN virus (SENV), and the proteins encoded by it. SENV is thought to be
 CC the cause of hepatopathies which are not linked to the presence of the
 CC hepatitis A, B and E viruses in man. The genome and proteins of this
 CC virus can be used in gene therapy and vaccination against the virus,
 CC which also causes disorders of the gastrointestinal tract, including
 CC Crohn's disease and lupus erythematosus, inflammatory diseases, and
 CC proliferative disorders such as cancer.
 CC
 XX
 SO Sequence 2040 BP; 692 A; 560 C; 387 G; 401 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 2040;
 Best Local Similarity 80.0%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggatcgtctgttgaagcagcagc 25
 ||| ||||| ||| |||||
 DB 1969 GGAGCTGCTGCTTGATGACGACCTG 1945

RESULT 36
 ID AAN80922 standard; CDNA; 2153 BP.
 AC AAN80922;
 XX
 DT 29-DEC-1990 (first entry)
 XX
 DE Sequence encoding human estrogen-related receptor protein (hERR2).
 XX
 KW Hormone receptor; hormone-binding; transcription activation; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FH CDS 100..1401
 FT /*tag= a
 XX
 PN W08803168-A.
 XX
 PD 05-MAY-1988.
 XX
 PF 23-OCT-1987; 87WO-US02782.
 XX
 PR 20-OCT-1987; 87US-0108471.

PR 04-OCT-1986; 86US-0922585.
XX
XX (SALK) SALK INST FOR BIOL STUD.
PA
XX Evans RM, Weinberger CA, Hollenberg SM, Giguere V;
PI WPI. 1988-133242/19.
XX
XX P-PSDB; AAB0931.
DR
XX Recombinant DNA encoding hormone receptors -
PT comprising glucocorticoid, mineralocorticoid, thyroid hormone
PT and novel hormone receptors
XX
XX Claim 16: Fig V-2(B)-1 and -2; 243pp; English.
PS
XX DNAs encoding hormone receptors and the hormone receptors themselves are
CC claimed. The DNA can be used to make the hormone receptor proteins and
CC functional modified forms in quantities not previously possible. The
CC receptor proteins can be used to screen cpds. for receptor-agonist or
CC receptor-antagonist activity. They can also be used in diagnostic assays.
CC Also claimed is pure DNA encoding oestrogen-related receptors hERR1 and
CC hERR2 and hERR1 and hERR2. The new hERR1 and hERR2 receptors will provide
CC the basis for development of an assay system that will lead to the
CC identification of novel hormones.
XX
XX Sequence 2153 BP; 445 A; 627 C; 624 G; 457 T; 0 other;
SQ

Query Match 68.0%; Score 17; DB 9; Length 2153;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggatctgctgtttgaagcgagcag 25
||| ||||| || ||||| |||||
DB 1200 ggacctgtcatgagcgagcgagcag 1224

RESULT 37
AA06819
ID AAX06819 standard; DNA; 2526 BP.
XX
XX AC AAX06819;
XX
XX DT 26-APR-1999 (first entry)
XX
XX DE Chlamydia pneumoniae surface exposed protein Omp7 DNA.
XX
XX KW Omp7; outer membrane protein 7; surface exposed protein; antigen;
KM infection; diagnosis; vaccine; atherosclerosis; asthma; ss.
XX
XX OS Chlamydia pneumoniae.
XX
XX PN WO9858953-A2.
XX
XX PD 30-DEC-1998.
XX
XX PF 19-JUN-1998; 98WO-DK00266.
XX
XX PR 23-JUN-1997; 97DK-0000744.
XX
XX PA (BIRK/) BIRKELUND S.
XX (CHRI/) CHRISTIANSEN G.
XX
XX PI Birkelund S, Christiansen G, Knudsen K, Madsen A;
PI Mylind P;
XX
XX DR WPI: 1999-105610/09.
XX P-PSDB; AAW88420.
XX
XX PT Species-specific test for identifying mammals infected with
PT Chlamydia pneumoniae - comprises detecting antibodies specific for
PT outer membrane proteins of C. pneumoniae or nucleic acids encoding
PT these proteins

XX
PS Claim 6; Page 49-50; 115pp; English.
XX
XX This DNA sequence codes for the novel 89.7 kDa surface exposed
CC protein Omp7 (see AAW88420) of the human respiratory pathogen
CC Chlamydia pneumoniae. By generating antibodies against C.
CC pneumoniae outer membrane complex, a polyclonal antibody (PAB 150)
CC was obtained which reacted with outer membrane proteins. The
CC antibody was used to identify the genes (see AAX06816-27) encoding
CC Omp4-Omp5 proteins (see AAW88417-28) in an expression library of
CC C. pneumoniae DNA. The genes are situated in 2 gene clusters:
CC Omp12,11,10,5,4,13 and 14 in one cluster and Omp6,7,8,9 and 15 in
CC the other, and encode polypeptides of about 89.6-100.3 kDa and
CC about 56.1 kDa. The invention provides a new species specific test
CC for identifying mammals (including humans) infected with Chlamydia
CC pneumoniae. The test comprises detecting antibodies specific for
CC Omp4-Omp5 or detecting nucleic acid fragments encoding these outer
CC membrane proteins, especially by PCR. The proteins are also used
CC in the diagnosis of C. pneumoniae infection in mammals. The
CC nucleic acids and proteins can also be used in the immunization of
CC mammals, the nucleic acids being particularly useful as DNA
CC vaccines for effecting in vivo expression of antigens. The
CC vaccines may also prevent atherosclerosis and bronchial asthma,
CC which are possibly associated with C. pneumoniae.
XX
XX Sequence 2526 BP; 696 A; 495 C; 597 G; 738 T; 0 other;
SQ

Query Match 68.0%; Score 17; DB 20; Length 2526;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggatctgctgtttgaagcgagcag 25
||| ||||| || ||||| |||||
DB 331 ggagctgctattggaagcgagcag 355

RESULT 38
AAA28691
ID AAA28691 standard; DNA; 2526 BP.
XX
XX AC AAA28691;
XX
XX DT 29-AUG-2000 (first entry)
XX
XX DE C. pneumoniae CPN100397 open reading frame.
XX
XX KW Antigen; anti-inflammatory; respiratory; antibacterial; anti-asthmatic;
KM anti-arteriosclerotic; vaccine; ds.
XX
XX OS Chlamydia pneumoniae.
XX
XX PN WO200024765-A2.
XX
XX PD 04-MAY-2000.
XX
XX PF 28-OCT-1999; 99WO-CA00992.
XX
XX PR 28-OCT-1998; 98US-0106034.
XX PR 28-OCT-1998; 98US-0106039.
XX PR 28-OCT-1998; 98US-0106042.
XX PR 28-OCT-1998; 98US-0106044.
XX PR 29-OCT-1998; 98US-0106072.
XX PR 29-OCT-1998; 98US-0106073.
XX PR 29-OCT-1998; 98US-0106074.
XX PR 29-OCT-1998; 98US-0106087.
XX PR 02-NOV-1998; 98US-0106587.
XX PR 02-NOV-1998; 98US-0106588.
XX PR 02-NOV-1998; 98US-0107034.
XX PR 02-NOV-1998; 98US-0107035.
XX
XX PA (CONN-) CONNAUGHT LAB LTD.
XX

PI Murdin AD, Oomen RP, Wang J;
XX
XX WPI: 2000-350688/30.
DR P-PSDB; AAY92818.
XX
XX Chlamydia antigens and the proteins they encode, useful for
PT vaccinating against Chlamydia infections that affect the respiratory
PT tract
XX
XX Claim 2: Fig 1: 226pp: English.
PS
XX
XX The nucleic acids may be used for the recombinant production of the
CC Chlamydia polypeptides (either in vivo or in vitro) according to standard
CC recombinant DNA methodologies. The polypeptides may then be used to
CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as
CC C. pneumoniae, are pathogens responsible for upper respiratory tract
CC infections such as community acquired pneumonia, acute respiratory
CC disease and bronchitis and may be implicated in atherosclerotic changes
CC and asthma. The nucleic acids may also be used as probes for detecting
CC the presence of Chlamydia nucleic acids in samples (and therefore for
CC diagnose infections) and the proteins may be used as antigens for the
CC production of antibodies that may be used to detect Chlamydia proteins
CC in samples (e.g. via enzyme linked immunosorbent assay (ELISA)).
XX
SQ Sequence 2526 BP; 693 A; 493 C; 601 G; 739 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 2526;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgagcag 25
||| ||||| ||| ||||| |||||
Db 331 ggaatcgtctgttgaagcgagcag 355

RESULT 39

AAA61153/C
ID AAA61153 standard; DNA: 2612 BP.

XX AAA61153;

XX 12-OCT-2000 (first entry)

XX SEN virus genome fragment SEQ ID NO: 94.

XX SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease;
KW proliferative disorder; hepatopathy; hepatitis; viral infection;
KW vaccination; gene therapy; ds.
XX
XX Hepatitis virus.
OS
XX
XX W0200028039-A2.
PN
XX 18-MAY-2000.
PD
XX
XX 09-NOV-1999; 99WO-EP08566.
PF
XX
XX 10-NOV-1998; 98IT-M102437.
PR
XX 30-APR-1999; 99IT-M100923.
PR
XX 14-MAY-1999; 99EP-0830298.
PR
XX 16-JUL-1999; 99EP-0113932.
PR
XX
XX (DIAS-) DIASORIN SRL.
PA
XX
XX Primi D, Fioralissi G, Mantero GL, Mattioli S, Sottini A;
PI Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;
XX
XX WPI: 2000-376551/32.
DR
XX
XX Nucleic acids representing the genome of the SEN virus (SENV) and
PT encoded proteins, useful for treatment of hepatopathies, inflammatory
PT diseases and proliferative disorders such as cancer -

XX
XX Claim 1: Page 344-345; 392pp: English.
PS
XX
XX The present invention is concerned with the sequence of the genome of the
CC SEN virus (SENV), and the proteins encoded by it. SENV is thought to be
CC the cause of hepatopathies which are not linked to the presence of the
CC hepatitis A, B and E viruses in man. The genome and proteins of this
CC virus can be used in gene therapy and vaccination against the virus,
CC which also causes disorders of the gastrointestinal tract, including
CC Crohn's disease and lupus erythematosus, inflammatory diseases, and
CC proliferative disorders such as cancer.
XX
SQ Sequence 2612 BP; 828 A; 736 C; 545 G; 503 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 2612;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgagcag 25
||| ||||| ||| ||||| |||||
Db 2541 GGACGTCTGTGATGACGACGCTG 2517

RESULT 40

AAA28690
ID AAA28690 standard; DNA: 2750 BP.

XX AAA28690;

XX 29-AUG-2000 (first entry)

XX C. pneumoniae CPN100397 gene.

XX Antigen; anti-inflammatory; respiratory; antibacterial; anti-asthmatic;
KW anti-arteriosclerotic; vaccine; ds.
KW Chlamydia pneumoniae.
XX
XX
XX Key Location/Qualifiers
FT CDS 101..2626
FT /*tag= a

XX W0200024765-A2.
PN
XX
XX 04-MAY-2000.
PD
XX
XX 28-OCT-1999; 99WO-CA00992.
PF
XX
XX 28-OCT-1998; 98US-0106034.
PR
XX 28-OCT-1998; 98US-0106039.
PR
XX 28-OCT-1998; 98US-0106042.
PR
XX 28-OCT-1998; 98US-0106044.
PR
XX 29-OCT-1998; 98US-0106072.
PR
XX 29-OCT-1998; 98US-0106073.
PR
XX 29-OCT-1998; 98US-0106074.
PR
XX 29-OCT-1998; 98US-0106087.
PR
XX 02-NOV-1998; 98US-0106587.
PR
XX 02-NOV-1998; 98US-0106588.
PR
XX 02-NOV-1998; 98US-0107034.
PR
XX 02-NOV-1998; 98US-0107035.
PR
XX
XX (CONN-) CONNAUGHT LAB LTD.
PA
XX
XX Murdin AD, Oomen RP, Wang J;
PI
XX
XX WPI: 2000-350688/30.
DR
XX
XX P-PSDB; AAY92818.
DR
XX
XX Chlamydia antigens and the proteins they encode, useful for
PT vaccinating against Chlamydia infections that affect the respiratory
PT tract
XX

PS Claim 2; Fig 1; 226pp; English.
XX
CC The nucleic acids may be used for the recombinant production of the
CC Chlamydia polypeptides (either in vivo or in vitro) according to standard
CC recombinant DNA methodologies. The polypeptides may then be used to
CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as
CC C. pneumoniae, are pathogens responsible for upper respiratory tract
CC infections such as community acquired pneumonia, acute respiratory
CC disease and bronchitis and may be implicated in atherosclerotic changes
CC and asthma. The nucleic acids may also be used as probes for detecting
CC the presence of Chlamydia nucleic acids in samples (and therefore
CC diagnose infections) and the proteins may be used as antigens for the
CC production of antibodies that may be used to detect Chlamydia proteins
CC in samples (e.g. via enzyme linked immunosorbent assay (ELISA)).
XX
SQ Sequence 2750 BP; 756 A; 532 C; 641 G; 821 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 2750;
Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gcatcgtctgttgaacgcagcag 25
||| ||||| || ||||| |||||
Db 431 ggaagctctctatggaagtgaagcag 455

RESULT 41
AAC46633/c
ID AAC46633 standard; DNA; 2895 BP.
XX AAC46633;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zee mays DNA fragment SEQ ID NO: 50852.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW Protein identification; signal transduction pathway; metabolic;
KW pathway; promoter; termination sequence; corn; ss.
XX
OS Zee mays subsp. mays.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132485.
PR 05-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137522.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

Query Match	Best Local	Similarity	Score 17;	DB 21;	Length 2895;
Matches	20;	Conservative	80.0%;	Pred. No. 1.7e+02;	
			0;	Mismatches 5;	Indels 0;
				Gaps	0;
QY	1	ggatctgctgtttgaagcgcagcag 25			
Db	2606	GGATCTCCTCGTGTGAAGGCTGAG 2582			
<p>RESULT 42</p> <p>AAA61126/c</p> <p>ID AAA61126 standard; DNA; 3619 BP.</p> <p>XX AAA61126;</p> <p>XX</p> <p>XX 12-OCT-2000 (first entry)</p> <p>XX</p> <p>DE SEN virus genome fragment SEQ ID NO: 34.</p> <p>XX</p> <p>XX SEN virus; SENV; gastrointestinal tract disorder; inflammatory disease;</p> <p>KW proliferative disorder; hepatopathy; hepatitis; viral infection;</p> <p>XX vaccination; gene therapy; ds.</p> <p>XX</p> <p>XX Hepatitis virus.</p> <p>XX</p> <p>OS WO200028039-A2.</p> <p>PN</p> <p>PD 18-MAY-2000.</p> <p>XX</p> <p>XX 09-NOV-1999; 99WO-EP08566.</p> <p>PE</p> <p>XX 10-NOV-1998; 98IT-MI02437.</p> <p>XX 30-APR-1999; 99IT-MI00923.</p> <p>PR 14-MAY-1999; 99EP-0830298.</p> <p>PR 16-JUL-1999; 99EP-0113932.</p> <p>XX</p> <p>FA (DIAS-) DIASORIN SRL.</p> <p>XX</p> <p>XX Primi D, Fioridalisi G, Mantero GL, Mattioli S, Sottini A;</p> <p>PI Bonelli F, Vaglini L, Olivero P, Dal Corso A, Bonelli M;</p> <p>PI WPI; 2000-376551/32.</p> <p>DR</p> <p>XX</p> <p>XX Nucleic acids representing the genome of the SEN virus (SENV) and</p> <p>PT encoded proteins, useful for treatment of hepatopathies, inflammatory</p> <p>PT diseases and proliferative disorders such as cancer -</p> <p>XX</p> <p>XX</p> <p>PS Claim 1; Page 311-312; 392pp; English.</p> <p>XX</p> <p>XX The present invention is concerned with the sequence of the genome of the</p> <p>CC SEN virus (SENV), and the proteins encoded by it. SENV is thought to be</p> <p>CC the cause of hepatopathies which are not linked to the presence of the</p> <p>CC hepatitis A, B and E viruses in man. The genome and proteins of this</p> <p>CC virus can be used in gene therapy and vaccination against the virus,</p> <p>CC which also causes disorders of the gastrointestinal tract, including</p> <p>CC Crohn's disease and lupus erythematosus, inflammatory diseases, and</p> <p>CC proliferative disorders such as cancer.</p> <p>XX</p> <p>XX Sequence 3619 BP; 1038 A; 1035 C; 825 G; 721 T; 0 other;</p>					

Best Local Similarity 80.0%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggcctcgtcgtttgaagcgagcag 25
||| ||||| ||||| ||||| ||||| |||||
Db 2800 GGAGCTGCTGCTGATGAGCAGCCTG 2776

RESULT 43

AAK91990
ID AAK91990 standard; DNA: 1230025 BP.

XX AC AAK91990;

XX DT 13-SEP-1999 (first entry)

XX DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

XX KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

XX KW vaccine; neutralising epitope; ss.

XX OS Chlamydia pneumoniae.

XX PN W09927105-A2.

XX PD 03-JUN-1999.

XX PF 20-NOV-1998; 98MO-IB01890.

XX PR 04-NOV-1998; 98US-0107078.

XX PR 21-NOV-1997; 97FR-0014673.

XX PA (GEST) GENSET.

XX PI Griffiths R;

XX XX WPI; 1999-357842/30.

XX PT Genome sequence of Chlamydia pneumoniae

XX PS Claim 1; Page 291-611; 1912pp; English.

XX CC The present sequence represents the complete genome of Chlamydia

XX CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes

XX CC respiratory disease such as pneumonia and bronchitis and is thought

XX CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,

XX CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides

XX CC encoded by the open reading frames of the C. pneumoniae genome (see

XX CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.

XX CC Vectors containing C. pneumoniae nucleotide sequences can also be

XX CC used as immunogenic compositions, especially where the vector directs

XX CC the expression of a neutralising epitope of C. pneumoniae.

XX CC Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

XX QY 1 ggcctcgtcgtttgaagcgagcag 25

XX DB 29280 ggaagctcgtcgtttgaagcgagcag 29304

XX RESULT 44
XX AAC06234/c
XX ID AAC06234 standard; CDNA: 365 BP.
XX AC AAC06234;
XX DT 06-OCT-2000 (first entry)

XX XX Human secreted protein 5' EST, SEQ ID NO: 10309.

XX DE Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX KW gene therapy; chromosome mapping; ss.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

XX PF 21-FEB-2000; 2000EP-0200610.

XX PR 26-FEB-1999; 99US-0122487.

XX PA (GEST) GENSET.

XX PI Dumas Mline Edwards J, Duclert A, Giordano J;

XX DR WPI; 2000-500381/45.

XX PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX PS Claim 1; SEQ ID 10309; 71pp + CD-ROM; English.

XX CC The present sequence is one of a large number of 5' ESTs derived from

XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively

XX CC identified within the present sequence. The 5' ESTs were prepared from

XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST

XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)

XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA

XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences

XX CC derived from the 5' ends of mRNAs and even in those cases where longer

XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.

XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used

XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX CC They are used to obtain upstream regulatory sequences and to design

XX CC expression and secretion vectors.

XX CC Sequence 365 BP; 115 A; 90 C; 90 G; 69 T; 1 other;

XX QY 6 tgctcgtttgaagcgagcag 25

XX DB 201 TGCTGTGTGAAGCGACGACAG 182

XX RESULT 45
XX AAC06233/c
XX ID AAC06233 standard; CDNA: 404 BP.

XX AC AAC06233;

XX DT 06-OCT-2000 (first entry)

XX DE Human secreted protein 5' EST, SEQ ID NO: 10308.

XX KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX KW gene therapy; chromosome mapping; ss.

XX OS Homo sapiens.

XX PN EP1033401-A2.

XX PD 06-SEP-2000.

```

XX 21-FEB-2000; 2000EP-0200610.
PF XX
XX 26-FEB-1999; 99US-0122487.
PR XX
XX (GENSET ) GENSET.
PA
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
DR
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 10308; 71pp + CD-ROM; English.
PS
XX
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SO Sequence 404 BP; 129 A; 96 C; 106 G; 73 T; 0 other;

```

```

Query Match 67.2%; Score 16.8; DB 21; Length 404;
Best Local Similarity 90.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 6 tactgttgaagcgcagcag 25
   ||||| ||||| |||||
DB 240 TCCTGTGTGAAGCGCACCA 221

```

Search completed: October 9, 2001, 11:37:32
 Job time: 1767 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: October 9, 2001, 11:39:37 ; Search time 221.13 Seconds
(without alignments)
21.403 Million cell updates/sec

Title: US-09-396-196f-4

Sequence: 1 ggatctgctgttgaagcgcagcag 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents, NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfilesl.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	2 US-08-401-068-7	Sequence 7, Appl1
2	25	100.0	1041	2 US-08-846-338-7	Sequence 7, Appl1
3	25	100.0	5872	3 US-08-411-7688-1	Sequence 1, Appl1
4	25	100.0	5872	3 US-08-411-7688-6	Sequence 6, Appl1
5	17.6	70.4	3561	1 US-08-097-997A-12	Sequence 12, Appl1
6	17.6	70.4	3561	3 US-08-665-574C-12	Sequence 12, Appl1
7	17.6	70.4	3561	4 US-08-946-994-12	Sequence 12, Appl1
8	17.6	70.4	15894	1 US-08-348-891A-1	Sequence 1, Appl1
9	17.6	70.4	15894	1 US-08-905-817-1	Sequence 1, Appl1
10	16.6	66.4	1835	1 US-08-361-611-3	Sequence 3, Appl1
11	16.6	66.4	1835	1 US-08-565-655-3	Sequence 3, Appl1
12	16.6	66.4	1835	2 US-08-946-967-3	Sequence 9, Appl1
13	16.2	64.8	1694	1 US-08-698-551-9	Sequence 9, Appl1
14	16.2	64.8	1694	2 US-08-602-228-9	Sequence 9, Appl1
15	16.2	64.8	1694	2 US-08-494-440B-9	Sequence 9, Appl1
16	16.2	64.8	1694	2 US-08-533-901B-9	Sequence 9, Appl1
17	16.2	64.8	1694	2 US-08-839-032A-9	Sequence 9, Appl1
18	16.2	64.8	1694	2 US-08-839-031A-9	Sequence 9, Appl1
19	16.2	64.8	1694	5 PCT-US95-12724-9	Sequence 9, Appl1
20	16.2	64.8	1810	1 US-07-755-573C-7	Sequence 7, Appl1
21	16.2	64.8	2473	1 US-08-698-551-17	Sequence 17, Appl1
22	16.2	64.8	2473	2 US-08-839-032A-17	Sequence 17, Appl1
23	16	64.0	1080	1 US-08-225-757B-1	Sequence 1, Appl1
24	16	64.0	1080	2 US-08-722-050-1	Sequence 1, Appl1
25	16	64.0	1248	3 US-08-910-505-3	Sequence 3, Appl1
26	16	64.0	1251	3 US-08-910-505-1	Sequence 1, Appl1
27	16	64.0	1467	1 US-08-176-620A-3	Sequence 3, Appl1

C 28	16	64.0	1467	1	US-08-463-862-3	Sequence 3, Appl1
C 29	16	64.0	1467	2	US-08-461-985-3	Sequence 3, Appl1
C 30	16	64.0	1467	2	US-08-458-887-3	Sequence 3, Appl1
C 31	16	64.0	1686	1	US-08-356-405-1	Sequence 1, Appl1
C 32	16	64.0	3147	4	US-09-101-886B-1	Sequence 11, Appl1
C 33	16	64.0	5430	3	US-09-012-515A-11	Sequence 11, Appl1
C 34	16	64.0	5430	3	US-08-360-144A-11	Sequence 11, Appl1
C 35	16	64.0	7824	5	PCT-US95-06722-11	Sequence 11, Appl1
C 36	16	64.0	11703	4	US-09-101-886B-3	Sequence 3, Appl1
C 37	16	64.0	13865	3	US-09-009-217-11	Sequence 11, Appl1
C 38	16	64.0	13865	3	US-09-009-656-11	Sequence 11, Appl1
C 39	15.8	63.2	1902	2	US-09-258-257-1	Sequence 1, Appl1
C 40	15.8	63.2	1902	2	US-09-258-371-1	Sequence 1, Appl1
C 41	15.8	63.2	1902	3	US-08-569-671A-1	Sequence 1, Appl1
C 42	15.8	63.2	1902	3	US-08-751-230-1	Sequence 1, Appl1
C 43	15.8	63.2	1902	3	US-09-499-082-1	Sequence 1, Appl1
C 44	15.8	63.2	1902	4	US-09-258-372-1	Sequence 1, Appl1
C 45	15.8	63.2	2061	3	US-09-499-082-9	Sequence 9, Appl1

ALIGNMENTS

RESULT 1
US-08-401-068-7
Sequence 7, Application US/08401068
Patent No. 5859335
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Enhanced Biotin Biotin synthesis in Plant Tissue
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/351,970
FILING DATE: 08-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

```
Query Match      100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0093;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ggatctgctgttgaagcgcagcag 25
        |||
Db       63  GGATCTGCTGTTGAAGCGCAGCAG 87

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5869719artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 36,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match      100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0093;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  ggatctgctgttgaagcgcagcag 25
        |||
Db       63  GGATCTGCTGTTGAAGCGCAGCAG 87

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: pBO30A-15/9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 117..1157
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 117
; OTHER INFORMATION: /product= "Biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bIob"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2295..3050
; OTHER INFORMATION: /codon_start= 2295
; OTHER INFORMATION: /function= "Involved in pimeloyl-CoA synthesis"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /gene= "bIoc"
; OTHER INFORMATION: /number= 3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3750..5039
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 3750
; OTHER INFORMATION: /EC_number= 2.6.1.62
; OTHER INFORMATION: /product= "DAPA synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioA"
; OTHER INFORMATION: /number= 5
; OTHER INFORMATION: /standard_name=
; OTHER INFORMATION: "S-adenosyl-L-methionine:8-amino-7-oxononanoate
; OTHER INFORMATION: aminotransf."
; FEATURE:
; NAME/KEY: CDS
```



```

? LOCATION: 5098..5574
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start= 5098
? OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
? OTHER INFORMATION: /product= "protein"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ORF1"
? OTHER INFORMATION: /number= 6
? FEATURE:
? NAME/KEY: -10_signal
? LOCATION: 45..49
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /standard_name= "promoter ptac"
? FEATURE:
? NAME/KEY: -35_signal
? LOCATION: 23..28
? OTHER INFORMATION: /standard_name= "promoter ptac"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 105..119
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /standard_name= "biob RBS no.9"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 2284..2297
? OTHER INFORMATION: /standard_name= "bioc RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 3742..3752
? OTHER INFORMATION: /standard_name= "biola RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 5088..5100
? OTHER INFORMATION: /standard_name= "ORF1 RBS"
? FEATURE:
? NAME/KEY: terminator
? LOCATION: 5583..5644
? OTHER INFORMATION: /standard_name= "rho-independent
? OTHER INFORMATION: /transcriptional terminator"
? FEATURE:
? NAME/KEY: stem_loop
? LOCATION: 5583..5605
? FEATURE:
? NAME/KEY: promoter
? LOCATION: 1..96
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /function= "promoter ptac"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? PUBLICATION INFORMATION:
? DOCUMENT NUMBER: WO 87/01391 B1
? FILING DATE: 26-AUG-1986
? PUBLICATION DATE: 07-APR-1993
? US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Oy 1 ggaatcgtcttgaagcgagcag 25
Db 179 ggaatcgtcttgaagcgagcag 203

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "biob RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biob RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
```

FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.013;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gactcgtctgttgaagcgcagcag 25
|||||
Db 179 GATCTGCTGCTTGAAGCGCAGCAG 203

RESULT 5
US-08-097-997A-12/C
Sequence 12, Application US/08097997A
Patent No. 5728536

GENERAL INFORMATION:
APPLICANT: Ihle, James N.
APPLICANT: Silvenoinen, Olli
APPLICANT: Wiltuhin, Bruce A.
APPLICANT: Quelle, Frederick W.
TITLE OF INVENTION: Jak Kinases and Regulation of Cytokine Signal
TITLE OF INVENTION: Transduction
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/097,997A
FILING DATE: 29-JULY-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fox, Samuel L.
REGISTRATION NUMBER: 30,353
REFERENCE/DOCKET NUMBER: 0656.0370000/SLF/GKT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3561
US-08-097-997A-12

Query Match 70.4%; Score 17.6; DB 1; Length 3561;
Best Local Similarity 83.3%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gactcgtctgttgaagcgcagcag 25
|||||
Db 3021 GATCTGCTGCGGAGAGCAGCAG 2998

RESULT 6
US-08-665-574C-12/C
Sequence 12, Application US/08665574C

Patent No. 6136595
GENERAL INFORMATION:
APPLICANT: Ihle, James N.
APPLICANT: Silvenoinen, Olli
APPLICANT: Wiltuhin, Bruce A.
TITLE OF INVENTION: Jak Kinases and Regulation of Cytokine
TITLE OF INVENTION: Signal Transduction
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3934

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,574C
FILING DATE: 18-JUN-1996
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/282,012
FILING DATE: 29-JUL-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/097,997
FILING DATE: 29-JUL-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/118,968
FILING DATE: 09-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 0656.0370002/SLF/LBB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3561
US-08-665-574C-12

Query Match 70.4%; Score 17.6; DB 3; Length 3561;
Best Local Similarity 83.3%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gactcgtctgttgaagcgcagcag 25
|||||
Db 3021 GATCTGCTGCGGAGAGCAGCAG 2998

RESULT 7
US-08-946-994-12/C
Sequence 12, Application US/08946994
Patent No. 6210654
GENERAL INFORMATION:
APPLICANT: Ihle, James N.
APPLICANT: Silvenoinen, Olli
APPLICANT: Wiltuhin, Bruce A.
APPLICANT: Quelle, Frederick W.
TITLE OF INVENTION: Jak Kinases and Regulation of Cytokine Signal
TITLE OF INVENTION: Transduction
NUMBER OF SEQUENCES: 17

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
;; STREET: 1100 New York Avenue, Suite 600
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20005-3934
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/946,994
;; FILING DATE:
;; CLASSIFICATION:
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US/08/665,574
;; FILING DATE: 18-JUN-1996
;; APPLICATION NUMBER: 08/282,012
;; FILING DATE: 29-JUL-1994
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: 08/097,997
;; FILING DATE: 29-JUL-1993
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: 08/118,968
;; FILING DATE: 09-SEP-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fox, Samuel L.
;; REGISTRATION NUMBER: 30,353
;; REFERENCE/DOCKET NUMBER: 0656.0370002/SLE/GKT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (202) 371-2600
;; TELEFAX: (202) 371-2540
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3561 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..3561
;;
US-08-946-994-12

Query Match 70.4%; Score 17.6; DB 4; Length 3561;
Best Local Similarity 83.3%; Pred. No. 26;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgacagcag 25
||||| ||||| ||||| ||||| |||||
DB 3021 GATCTGCTGTGAGCAAGCAGCAG 2998

RESULT 8
US-08-348-891A-1
; Sequence 1, Application US/08348891A
; Patent No. 5654136
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Keiko
; APPLICANT: MORI, Takayuki
; APPLICANT: MAKINO, Satoshi
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
; TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA

;; ZIP: 22202
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/348,891A
;; FILING DATE: 25-NOV-1994
;; CLASSIFICATION: 435
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 07/848,400
;; FILING DATE: 10-MAR-1992
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: JP 3-293625
;; FILING DATE: 14-OCT-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: PATCH, Andrew J.
;; REGISTRATION NUMBER: 32,925
;; REFERENCE/DOCKET NUMBER: KP-7501
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-521-2297
;; TELEFAX: 703-685-0573
;; TELEX: 248425 EMBON
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 15894 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 108..1682
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1807..3327
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3438..4442
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 5458..7107
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 7271..9121
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 9234..15782
;;
US-08-348-891A-1

Query Match 70.4%; Score 17.6; DB 1; Length 15894;
Best Local Similarity 83.3%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgacagcag 25
||||| ||||| ||||| ||||| |||||
DB 12940 GATCTGCTGTGAGCAAGCAGCAG 12963

RESULT 9
US-08-905-817-1
; Sequence 1, Application US/08905817
; Patent No. 5824777
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Keiko
; APPLICANT: MORI, Takayuki
; APPLICANT: MAKINO, Satoshi
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
; TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
; NUMBER OF SEQUENCES: 19

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: YOUNG & THOMPSON
;; STREET: 745 South 23rd Street
;; CITY: Arlington
;; STATE: Virginia
;; COUNTRY: USA
;; ZIP: 22202
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent Release #1.0, Version #1.30
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/905,817
;; FILING DATE: 04-AUG-1997
;; CLASSIFICATION: 435
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/348,891
;; FILING DATE: 25-NOV-1994
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/848,400
;; FILING DATE: 10-MAR-1992
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 3-293625
;; FILING DATE: 14-OCT-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: PATCH, Andrew J
;; REGISTRATION NUMBER: 32,925
;; REFERENCE/DOCKET NUMBER: KP-7501A
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-521-2297
;; TELEFAX: 703-685-0573
;; TELEX: 248425 EMBON
;;
;; INFORMATION FOR SEQ ID NO: 1:
;;
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 15894 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;;
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 108..1682
;;
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1807..3327
;;
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 3438..4442
;;
;; NAME/KEY: CDS
;; LOCATION: 5458..7107
;;
;; NAME/KEY: CDS
;; LOCATION: 7271..9121
;;
;; NAME/KEY: CDS
;; LOCATION: 9234..15782
;;
;; US-08-905-817-1
;;
Query Match 70.4%; Score 17.6; DB 1; Length 15894;
Best Local Similarity 83.3%; Pred. No. 33;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctgtgttgaagcagcag 25
|||||

RESULT 10
US-08-361-611-3
; Sequence 3, Application US/08361611

;; Patent No. 5519125
;;
;; GENERAL INFORMATION:
;; APPLICANT: Potter, Sharon L
;; TITLE OF INVENTION: Plant Adenylosuccinate Synthetase and
;; NUMBER OF SEQUENCES: 4
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CIBA-GEIGY Corporation
;; STREET: 7 Skyline Drive
;; CITY: Hawthorne
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 10532
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent Release #1.0, Version #1.30B
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/361,611
;; FILING DATE:
;; CLASSIFICATION: 536
;;
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Elmer, James Scott
;; REGISTRATION NUMBER: 36,129
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (919) 541-8614
;; TELEFAX: (919) 541-8689
;;
;; INFORMATION FOR SEQ ID NO: 3:
;;
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1835 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; HYPOTHETICAL: NO
;;
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 18..1469
;;
;; OTHER INFORMATION:
;; OTHER INFORMATION: /product= "Maize Adenylosuccinate Synthetase"
;;
US-08-361-611-3

Query Match 66.4%; Score 16.6; DB 1; Length 1835;
Best Local Similarity 82.6%; Pred. No. 66;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 atctgtgttgaagcagcag 25
|||||

DB 508 ATCTGCTGTTGATCTGCACAG 530

RESULT 11
US-08-565-655-3

; Sequence 3, Application US/08565655

; Patent No. 568939

;; GENERAL INFORMATION:
;; APPLICANT: Potter, Sharon L
;; TITLE OF INVENTION: Plant Adenylosuccinate Synthetase and
;; NUMBER OF SEQUENCES: 6
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Ciba Patent Department
;; STREET: 540 White Plains Rd., POB 2005
;; CITY: Tarrytown
;; STATE: NY
;; COUNTRY: USA
;; ZIP: 10591-9005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/565,655
FILING DATE:
CLASSIFICATION: 210
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/361,611
FILING DATE: 12-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1835 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 18..1469
OTHER INFORMATION: /product= "Maize Adenylosuccinate synthetase"
US-08-565-655-3

Query Match 66.4%; Score 16.6; DB 1; Length 1835;
Best Local Similarity 82.6%; Pred. No. 66;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 atctgctgttgaagcgacag 25
|||||
DB 508 ATCTGCTGTGATCTGCACAG 530

RESULT 12
US-08-946-967-3
Sequence 3, Application US/08946967
Patent No. 5882869
GENERAL INFORMATION:
APPLICANT: Potter, Sharon L
APPLICANT: Ward, Eric R
TITLE OF INVENTION: Plant Adenylosuccinate Synthetase and
TITLE OF INVENTION: DNA Coding Therefor
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,967
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919) 541-8614
TELEFAX: (919) 541-8689

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1835 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 18..1469
OTHER INFORMATION: /product= "Maize Adenylosuccinate synthetase"
US-08-946-967-3

Query Match 66.4%; Score 16.6; DB 2; Length 1835;
Best Local Similarity 82.6%; Pred. No. 66;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 atctgctgttgaagcgacag 25
|||||
DB 508 ATCTGCTGTGATCTGCACAG 530

RESULT 13
US-08-698-551-9/c
Sequence 9, Application US/08698551
Patent No. 5712381
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
APPLICANT: Graham, James
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/698,551
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232D
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-698-551-9

Query Match 64.8%; Score 16.2; DB 1; Length 1694;
Best Local Similarity 85.7%; Pred. No. 98;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgagcag 25
||||| |||||||||
DB 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 14

US-08-602-228-9/c
; Sequence 9, Application US/086022228
; Patent No. 5843675
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/602,228
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A,
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: G15232C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1694 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..931
; US-08-602-228-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 98;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgagcag 25
||||| |||||||||
DB 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 15

US-08-494-440B-9/c
; Sequence 9, Application US/08494440B
; Patent No. 5849501
; GENERAL INFORMATION:

APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schievella, Andrea
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/494,440B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-494-440B-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 98;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgagcag 25
||||| |||||||||
DB 754 CTGCTGCCAGAGCGCAGCAG 734

RESULT 16

US-08-533-901B-9/c
; Sequence 9, Application US/08533901B
; Patent No. 5852173
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/533,901B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-533-901B-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 98;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgacagcag 25
||||| |||||||||
Db 754 CTGCTGCCAGAGCGACGACAG 734

RESULT 17
US-08-839-032A-9/c
Sequence 9, Application US/08839032A
Patent No. 5891675
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schlievela, Andrea
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,032A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A,
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232DDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:

LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-839-032A-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 98;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgacagcag 25
||||| |||||||||
Db 754 CTGCTGCCAGAGCGACGACAG 734

RESULT 18
US-08-839-031A-9/c
Sequence 9, Application US/08839031A
Patent No. 5948638
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schlievela, Andrea
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,031A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.,
REGISTRATION NUMBER: 41,323
REFERENCE/DOCKET NUMBER: G15232BDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 2..931
US-08-839-031A-9

Query Match 64.8%; Score 16.2; DB 2; Length 1694;
Best Local Similarity 85.7%; Pred. No. 98;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgctgttgaagcgagcag 25
||||| |||||||||
Db 754 CTGCTGCAGAGCGCAGCAG 734

RESULT 19
PCT-US95-12724-9/c
; Sequence 9, Application PC/TUS9512724
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/12724
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A., 32.724
; REGISTRATION NUMBER: 615232B
; REFERENCE/DOCKET NUMBER: 615232B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1694 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHEICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2..931
; PCT-US95-12724-9

Query Match 64.8%; Score 16.2; DB 5; Length 1694;
Best Local Similarity 85.7%; Pred. No. 98;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgctgttgaagcgagcag 25
||||| |||||||||
Db 754 CTGCTGCAGAGCGCAGCAG 734

RESULT 20
US-07-755-573C-7
; Sequence 7, Application US/07755573C
; Patent No. 5622930
; GENERAL INFORMATION:
; APPLICANT: Elderling, Eric
; APPLICANT: Arden, Lucien
; TITLE OF INVENTION: Cl Esterase Inhibitor Mutelins
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/755,573C
; FILING DATE: 05-SEP-1991
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Pochopien, Donald J.
; REGISTRATION NUMBER: 32,167
; REFERENCE/DOCKET NUMBER: 28687/32920
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1810 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..1535
; FEATURE:
; NAME/KEY: mat.peptide
; LOCATION: 102..1535
; US-07-755-573C-7

Query Match 64.8%; Score 16.2; DB 1; Length 1810;
Best Local Similarity 85.7%; Pred. No. 99;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 ctgctgttgaagcgagcag 25
||||| |||||||||
Db 1440 CTGCTCTTGAAGTCGAGCAG 1460

RESULT 21
US-08-698-551-17/c
; Sequence 17, Application US/08698551
; Patent No. 5712381
; GENERAL INFORMATION:
; APPLICANT: Lin, Lih-Ling
; APPLICANT: Chen, Jennifer H.
; APPLICANT: Schievella, Andrea
; APPLICANT: Graham, James
; TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND
; TITLE OF INVENTION: PROTEINS AND INHIBITORS OF LIGAND BINDING
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

FEATURE:
NAME/KEY: CDS
LOCATION: 14..2404
US-08-839-032A-17

Query Match 64.8%; Score 16.2; DB 1; Length 2473;
Best Local Similarity 85.7%; Pred. No. 1e+02; 3; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 3;

Db 2227 CTGCTGCCAGAGCGCAGCAG 2207

RESULT 22
US-08-839-032A-17/c
Sequence 17, Application US/08839032A
Patent No. 5891675
GENERAL INFORMATION:
APPLICANT: Lin, Lih-Ling
APPLICANT: Chen, Jennifer H.
APPLICANT: Schiavella, Andrea
TITLE OF INVENTION: NOVEL TNF RECEPTOR DEATH DOMAIN LIGAND PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/839,032A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Brown, Scott A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: G15232DDIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2473 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA

FEATURE:
NAME/KEY: CDS
LOCATION: 14..2404
US-08-839-032A-17

Query Match 64.8%; Score 16.2; DB 2; Length 2473;
Best Local Similarity 85.7%; Pred. No. 1e+02; 3; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 3;

Db 2227 CTGCTGCCAGAGCGCAGCAG 2207

RESULT 23
US-08-225-757B-1/c
Sequence 1, Application US/08225757B
Patent No. 5506133
GENERAL INFORMATION:
APPLICANT: YU, ET AL.
TITLE OF INVENTION: Superoxide Dismutase-4
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLIAN,
ADDRESSEE: CECCHI, STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NEW JERSEY
COUNTRY: USA
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH DISKETTE
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WORD PERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/225,757B
FILING DATE: 11 APR-94
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-106
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1080 BASE PAIRS
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA
US-08-225-757B-1

Query Match 64.0%; Score 16; DB 1; Length 1080;
Best Local Similarity 79.2%; Pred. No. 1e+02; 5; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 5;

Db 385 GGTCAGCTGTAGGAAGCGACCA 362

RESULT 24
US-08-722-050-1/c
Sequence 1, Application US/08722050
Patent No. 5871729

```

; GENERAL INFORMATION:
; APPLICANT: YU, GUO-LIANG
; APPLICANT: ROSEN, CRAIG A.
; APPLICANT: FRASER, CLAIRE M.
; APPLICANT: GOCAYNE, JEANNINE D.
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE-4
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,050
; FILING DATE: 23-JAN-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/225,757
; FILING DATE: 11-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: STREEPE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1020001/EKS/AJK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1080 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (cDNA)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 115..879
; US-08-722-050-1

Query Match          64.0%; Score 16; DB 2; Length 1080;
Best Local Similarity 79.2%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgacga 24
    ||| ||| ||| ||| ||| ||| |||
DB 385 GGTCAGCTGTGAGGAGCGACGA 362

RESULT 25
; US-08-910-505-3/c
; Sequence 3, Application US/08910505A
; Patent No. 6107071
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
; APPLICANT: Reichard, Raymond
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; APPLICANT: Lonetto, Michael
; APPLICANT: Warren, Patrick
; TITLE OF INVENTION: NOVEL HISTIDINOL DEHYDROG ENASE
; CORRESPONDENCE ADDRESS:
; FILE REFERENCE: P50549-02
```

```

; CURRENT APPLICATION NUMBER: US/08/910,505A
; CURRENT FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1248
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-08-910-505-3

Query Match          64.0%; Score 16; DB 3; Length 1248;
Best Local Similarity 79.2%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgacga 24
    ||| ||| ||| ||| ||| ||| |||
DB 968 GGATTACTGTTGAATGACGACGA 945

RESULT 26
; US-08-910-505-1/c
; Sequence 1, Application US/08910505A
; Patent No. 6107071
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie
; APPLICANT: Reichard, Raymond
; APPLICANT: Rosenberg, Martin
; APPLICANT: Ward, Judith
; APPLICANT: Lonetto, Michael
; APPLICANT: Warren, Patrick
; TITLE OF INVENTION: NOVEL HISTIDINOL DEHYDROG ENASE
; FILE REFERENCE: P50549-02
; CURRENT APPLICATION NUMBER: US/08/910,505A
; CURRENT FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1251
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-08-910-505-1

Query Match          64.0%; Score 16; DB 3; Length 1251;
Best Local Similarity 79.2%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgacga 24
    ||| ||| ||| ||| ||| ||| |||
DB 968 GGATTACTGTTGAATGACGACGA 945

RESULT 27
; US-08-176-620A-3/c
; Sequence 3, Application US/08176620A
; Patent No. 5595904
; GENERAL INFORMATION:
; APPLICANT: Boulton, Teri G.
; APPLICANT: Cobb, Melanie H.
; APPLICANT: Yancopoulos, George D.
; APPLICANT: Nye, Steven
; APPLICANT: Panayiotos, Nikos
; TITLE OF INVENTION: A Family of Map2 Protein Kinases
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
```

```

? CITY: New York
? STATE: New York
? COUNTRY: U.S.A.
? ZIP: 10036
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/176,620A
? FILING DATE: 03-JAN-1994
? CLASSIFICATION: 800
? ATTORNEY/AGENT INFORMATION:
? NAME: Mistock, S. Leslie
? REGISTRATION NUMBER: 18,872
? REFERENCE/DOCKET NUMBER: 6526-123
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (212) 790-9090
? TELEFAX: (212) 869-8864/9741
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1467 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: unknown
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..1245
? US-08-176-620A-3

Query Match          64.0%; Score 16; DB 1; Length 1467;
Best Local Similarity 79.2%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgagcag 25
    1 111 111 111111111 11
Db 411 GTCTCATGTCTGAAGCGCAGTAG 388

RESULT 28
US-08-463-862-3/c
? Sequence 3, Application US/08463862
? Patent No. 5776751
? GENERAL INFORMATION:
? APPLICANT: Boulton, Teri G. et al.
? TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
? NUMBER OF SEQUENCES: 11
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Pennie & Edmonds
? STREET: 1155 Avenue of the Americas
? CITY: New York
? STATE: New York
? COUNTRY: U.S.A.
? ZIP: 10036-2711
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/463,862
? FILING DATE: 05-JUN-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 07/701,544
? FILING DATE: 16-MAY-1991
? ATTORNEY/AGENT INFORMATION:
? NAME: Mistock, S. Leslie
? REGISTRATION NUMBER: 18,872
```

```

? REFERENCE/DOCKET NUMBER: 6526-049
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212 790-9090
? TELEFAX: 212 869-8864/9741
? TELE: 66141 PENNIE
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1467 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: unknown
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..1245
? US-08-463-862-3

Query Match          64.0%; Score 16; DB 1; Length 1467;
Best Local Similarity 79.2%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgagcag 25
    1 111 111 111111111 11
Db 411 GTCTCATGTCTGAAGCGCAGTAG 388

RESULT 29
US-08-461-985-3/c
? Sequence 3, Application US/08461985
? Patent No. 5872006
? GENERAL INFORMATION:
? APPLICANT: Boulton, Teri G.
? APPLICANT: Cobb, Melanie H.
? APPLICANT: Yancopoulos, George D.
? APPLICANT: Nye, Steven
? APPLICANT: Panayiotatos, Nikos
? TITLE OF INVENTION: A Family of Map2 Protein Kinases
? NUMBER OF SEQUENCES: 21
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Pennie & Edmonds
? STREET: 1155 Avenue of the Americas
? CITY: New York
? STATE: New York
? COUNTRY: U.S.A.
? ZIP: 10036
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: IBM PC compatible
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/461,985
? FILING DATE: 05-JUN-1995
? CLASSIFICATION: 800
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/176,620
? FILING DATE: 03-JAN-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Mistock, S. Leslie
? REGISTRATION NUMBER: 18,872
? REFERENCE/DOCKET NUMBER: 6526-123
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (212) 790-9090
? TELEFAX: (212) 869-8864/9741
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1467 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: unknown
? MOLECULE TYPE: DNA (genomic)
```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1245
US-08-461-985-3

```

Query Match	64.0%;	Score 16;	DB 2;	Length 1467;
Best Local Similarity	79.2%;	Pred. No. 1.2e+02;		
Matches 19; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;

QY	2	g	a	t	c	t	g	c	t	g	t	t	t	a	a	g	c	a	c	a	g	25
Db	411	G	T	T	C	A	T	G	T	C	T	G	A	G	C	C	A	G	T	A	388	

RESULT 30
US-08-458-887-3/c

GENERAL INFORMATION:
APPLICANT: Boulton, Teri G. et al.
TITLE OF INVENTION: A FAMILY OF MAP2 PROTEIN KINASES
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA

Query Match	64.0%;	Score 16;	DB 2;	Length 1467;
Best Local Similarity	79.2%;	Pred. No. 1.2e+02;		
Matches 19;	Conservative	0;	Mismatches 5;	Indels 0;
				Gaps 0;
QY	2	gacatgctgtttgaagcagcagcag	25	
Db	411	gTTCCTCATGTCTCTCAAGCGCAGTAG	388	

US-08-356-405-1
; Sequence 1, Application US/08356405
; Patent No. 5807691
; GENERAL INFORMATION:
ADDITONAL INFORMATION:

```

1 2:P: 19426-0107
2
3  COMPUTER READABLE FORM:
4
5  MEDIUM TYPE: Floppy disk
6
7  COMPUTER: IBM PC compatible
8
9  OPERATING SYSTEM: PC DOS/MS-DOS
10
11 SOFTWARE: Patentin Release #1.0, Version #1.25
12
13 CURRENT APPLICATION DATA:

```

Query Match	64.0%	Score 16;	DB 1;	Length 1686;
Best Local Similarity	79.2%;	Pred. No. 1.2e+02;		
Matches 19;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0

QY	1	g	a	t	c	t	c	t	t	t	a	a	c	c	a	c	a	24
Db	317	G	G	C	T	G	C	T	G	C	T	G	T	A	G	C	C	340

```

RESULT      32
US-09-101-8868-1/C
; Sequence 1, Application US/09101886B
; Patent No. 6197507
;
; GENERAL INFORMATION:
;
; APPLICANT:  BERG, THOMAS
; APPLICANT:  TOLLERSRUD, OLE K
; APPLICANT:  NILSSON, OIVIND
; TITLE OF INVENTION:  GENETIC TEST FOR ALPHA-MANNOSIDOSIS
;
; NUMBER OF SEQUENCES:  104

```



```

; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5430 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..5427
; US-08-360-144A-11

Query Match
Best Local Similarity 64.0%; Score 16; DB 3; Length 5430;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgctgttgaagcagcagca 24
    ||||| ||||| ||||| |||||
Db 3205 GGTTCGTGATGTTGTAGTGTACCA 3182

RESULT 35
PCT-US95-06722-11/c
; Sequence 11, Application PC/TUS9506722
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Immunosuppressant Target Proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06722
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/250,795
; FILING DATE: 27-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/250,795
; FILING DATE: 20-DEC-1994
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 97..7743
; PCT-US95-06722-11

Query Match
Best Local Similarity 64.0%; Score 16; DB 5; Length 7824;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggaatcgctgttgaagcagcagca 24
    ||||| ||||| ||||| |||||
Db 5521 GGTTCGTGATGTTGTAGTGTACCA 5498

RESULT 36
US-09-101-886B-3/c
; Sequence 3, Application US/09101886B
; Patent No. 6197507
; GENERAL INFORMATION:
; APPLICANT: BERG, THOMAS
; APPLICANT: TOLLERSRUD, OLE K
```

```

; APPLICANT: NILSEN, OIVIND
; TITLE OF INVENTION: GENETIC TEST FOR ALPHA-MANNOSIDOSIS
; NUMBER OF SEQUENCES: 104
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BARBARA G. ERNST
; STREET: 555 13TH STREET, NW SUITE 701E
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20004

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/101,886B
; FILING DATE: 29-JANUARY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB97/00109
; FILING DATE: 12-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: ERNST, BARBARA G
; REGISTRATION NUMBER: 30,377
; REFERENCE/DOCKET NUMBER: 1181-240
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-783-6040
; TELEFAX: 202-783-6031
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11703 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-09-101-886B-3

Query Match
Best Local Similarity 64.0%; Score 16; DB 4; Length 11703;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcagcagca 25
    ||||| ||||| ||||| |||||
Db 10828 GAACGTGCTCTCAAGCGCAGCAG 10805

RESULT 37
US-09-009-217-11/c
; Sequence 11, Application US/09009217
; Patent No. 6132729
; GENERAL INFORMATION:
; APPLICANT: Thorpe, Philip E.
; APPLICANT: King, Steven W.
; TITLE OF INVENTION: COMBINED TISSUE FACTOR AND
; TITLE OF INVENTION: CHEMOTHERAPEUTIC METHODS AND COMPOSITIONS FOR COAGULATION
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

FILING DATE: 27-JAN-1997

NAME/KEY: CDS

LOCATION: 109..738
US-09-258-257-1

Query Match 63.2%; Score 15.8; DB 2; Length 1902;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 gctgttgaagcgacag 25
|||||
Db 234 GCTGTTCGAGCGCAGCAG 252

RESULT 40
US-09-258-371-1
; Sequence 1, Application US/09258371
; Patent No. 5986078
; GENERAL INFORMATION:

APPLICANT: Garkavtsev, Igor
APPLICANT: Riabowol, Karl
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: Burns, Doane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA

COUNTRY: USA
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/258,371
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/751,230
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Mooi, Leslie A.
REGISTRATION NUMBER: 37,047

REFERENCE/DOCKET NUMBER: 028722-144
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-7400
TELEFAX: 415-854-8275

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1902 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:

NAME/KEY: CDS
LOCATION: 109..741
US-09-258-371-1

Query Match 63.2%; Score 15.8; DB 2; Length 1902;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 gctgttgaagcgacag 25
|||||
Db 234 GCTGTTCGAGCGCAGCAG 252

RESULT 41
US-08-569-721A-1

Sequence 1, Application US/08569721A
Patent No. 6037121
GENERAL INFORMATION:

APPLICANT: GARKAVTSEV, Igor
APPLICANT: RIABOWOL, Karl
TITLE OF INVENTION: DNA SEQUENCE ENCODING A TUMOR
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,721A
FILING DATE: 08-DEC-1995
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:
NAME: Mooi, Leslie A.
REGISTRATION NUMBER: 37,047

REFERENCE/DOCKET NUMBER: 028722-128
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 854-7400
TELEFAX: (650) 854-8275

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1902 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:

NAME/KEY: CDS
LOCATION: 109..738
US-08-569-721A-1

Query Match 63.2%; Score 15.8; DB 3; Length 1902;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 gctgttgaagcgacag 25
|||||
Db 234 GCTGTTCGAGCGCAGCAG 252

RESULT 42
US-08-751-230-1
; Sequence 1, Application US/08751230
; Patent No. 6117633
; GENERAL INFORMATION:

APPLICANT: Garkavtsev, Igor
APPLICANT: Riabowol, Karl
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: Burns, Doane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

Query Match 63.2%; Score 15.8; DB 3; Length 1902;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 7 gctgttgaagcgacag 25
|||||
Db 234 GCTGTTCGAGCGCAGCAG 252

RESULT 42
US-08-751-230-1
; Sequence 1, Application US/08751230
; Patent No. 6117633
; GENERAL INFORMATION:

APPLICANT: Garkavtsev, Igor
APPLICANT: Riabowol, Karl
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: Burns, Doane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/751,230
FILING DATE: 15-NOV-1996
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/569721
FILING DATE: 08-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Mooi, Leslie A.
REGISTRATION NUMBER: 37,047
REFERENCE/DOCKET NUMBER: 028722-144
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-7400
TELEFAX: 415-854-8275
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1902 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 109..741
US-08-751-230-1

Query Match 63.2%; Score 15.8; DB 3; Length 1902;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 7 gctgttgaagcgcagcag 25
||||| |||||||
DB 234 GCTGTTGAGGCGCAGCAG 252

RESULT 43
US-09-499-082-1
Sequence 1, Application US/09499082
Patent No. 6143522
GENERAL INFORMATION:
APPLICANT: Helbing, Karen C.
APPLICANT: Rlabowol, Karl
APPLICANT: Johnston, Randall N.
APPLICANT: Garkavtsev, Igor
TITLE OF INVENTION: METHODS OF MODULATING APOPTOSIS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/499,082
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/828,158
FILING DATE: 27-MAR-1997
APPLICATION NUMBER: US 08/751230
FILING DATE: 15-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/569721
FILING DATE: 08-DEC-1995

ATTORNEY/AGENT INFORMATION:
NAME: Mooi, Leslie A.
REGISTRATION NUMBER: 37,047
REFERENCE/DOCKET NUMBER: 028722-148
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-854-7400
TELEFAX: 650-854-8275
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1902 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 109..741
US-09-499-082-1

Query Match 63.2%; Score 15.8; DB 3; Length 1902;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 7 gctgttgaagcgcagcag 25
||||| |||||||
DB 234 GCTGTTGAGGCGCAGCAG 252

RESULT 44
US-09-258-372-1
Sequence 1, Application US/09258372
Patent No. 6238918
GENERAL INFORMATION:
APPLICANT: Garkavtsev, Igor
APPLICANT: Rlabowol, Karl
TITLE OF INVENTION: DNA SEQUENCE ENCODING THE TUMOR
SUPPRESSOR GENE INGI
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/258,372
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/751,230
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Mooi, Leslie A.
REGISTRATION NUMBER: 37,047
REFERENCE/DOCKET NUMBER: 028722-144
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-7400
TELEFAX: 415-854-8275
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1902 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:

NAME/KEY: CDS
LOCATION: 109..741
US-09-258-372-1

Query Match 63.2%; Score 15.8; DB 4; Length 1902;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 7 gctgttgaagcgagcag 25
|||||
Db 234 GCTGTCGAGCGCAGCAG 252

QY 7 gctgttgaagcgagcag 25
|||||
Db 393 GCTGTCGAGCGCAGCAG 411
Search completed: October 9, 2001, 11:39:41
Job time: 1896 sec

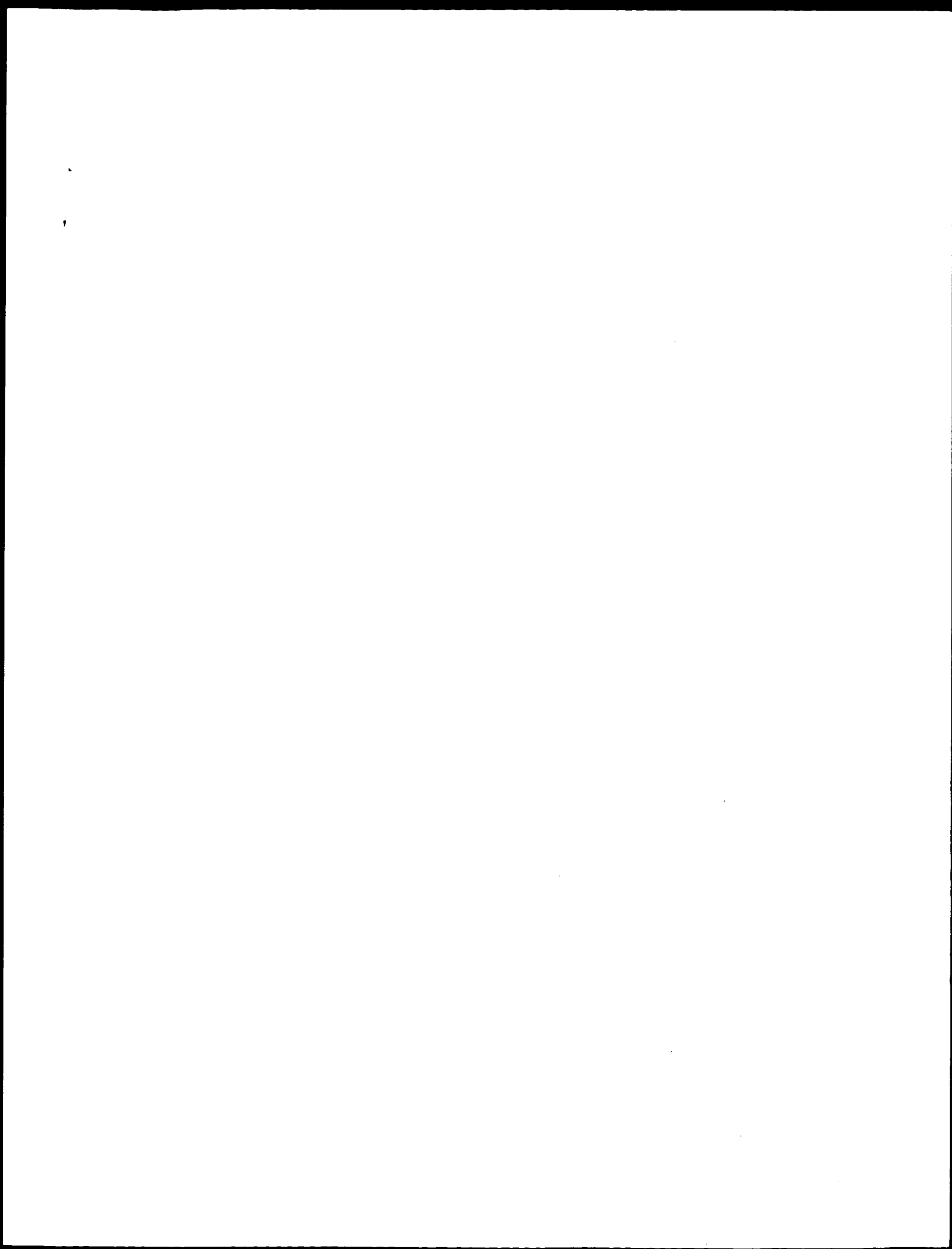
RESULT 45
US-09-499-082-9
Sequence 9, Application US/09499082
Patent No. 6143522
GENERAL INFORMATION:
APPLICANT: Helbing, Caren C.
APPLICANT: Riadowol, Karl
APPLICANT: Johnston, Randall N.
APPLICANT: Garikavtsev, Igor
TITLE OF INVENTION: METHODS OF MODULATING APOPTOSIS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/499,082
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/828,158
FILING DATE: 27-MAR-1997
APPLICATION NUMBER: US 08/751230
FILING DATE: 15-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/569721
FILING DATE: 08-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Mool, Leslie A.
REGISTRATION NUMBER: 37,047
REFERENCE/DOCKET NUMBER: 028722-148
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-854-7400
TELEFAX: 650-854-8275
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2061 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 16..900
US-09-499-082-9

Query Match 63.2%; Score 15.8; DB 3; Length 2061;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Wed Oct 10 07:44:36 2001

us-09-396-196f-4.std.rni

Page 21



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 13:46:33 ; Search time 5930.9 Seconds
(without alignments)
39.846 Million cell updates/sec

Title: US-09-396-196F-4

Perfect score: 25

Sequence: 1 ggatcgtcgttgaagcgacgacg 25

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1022815 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est110:*
191: gb_est111:*
192: gb_hlc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	19.8	79.2	642	21	A1514166	A1514166 GH27238.5
2	19.4	77.6	507	163	BE114310	BE114310 UI-R-BJ1-
3	19.4	77.6	507	163	AA818607	AA818607 UI-R-A0-b
4	19.2	76.8	437	229	AQ522448	AQ522448 HS.5237_A
5	19.2	76.8	1043	222	CNS054VD	AL3321250 Tetradon
6	18.8	75.2	238	26	AV231852	AV231852 Tetradon
7	18.8	75.2	520	122	AM944633	AM944633 SD10401.3
8	18.8	75.2	575	115	AM397892	AM397892 sg99e12.y
9	18.6	74.4	802	220	CNS02Y61	AL219267 Tetradon
10	18.6	74.4	466	104	AJ394746	AJ394746 AJ394746
11	18.6	74.4	468	32	AV680177	AV680177 AV680177
12	18.6	74.4	509	32	AV677254	AV677254 AV677254
13	18.6	74.4	617	32	AV675262	AV675262 AV675262
14	18.6	74.4	904	221	CNS04A1A	AL281899 Tetradon
15	18.6	74.4	930	221	CNS03FNU	AL241912 Tetradon
16	18.4	73.6	610	222	CNS051ED	AL316750 Tetradon
17	18.2	72.8	185	25	AV142481	AV142481 AV142481
18	18.2	72.8	243	11	AA750666	AA750666 TSGS0248
19	18.2	72.8	231	28	AV282092	AV282092 AV282092
20	18.2	72.8	268	134	BB452775	BB452775 BB452775
21	18.2	72.8	269	26	AV216567	AV216567 AV216567
22	18.2	72.8	283	25	AV136669	AV136669 AV136669
23	18.2	72.8	288	26	AV214322	AV214322 AV214322
24	18.2	72.8	319	25	AV169767	AV169767 AV169767
25	18.2	72.8	461	30	AV422424	AV422424 AV422424
26	18.2	72.8	584	174	BE160914	BE160914 de38c07.x
27	18.2	72.8	685	18	A1325366	A1325366 m129f01.y
28	18.2	72.8	765	108	AU170196	AU170196 AU170196
29	18.2	72.8	931	173	BE107933	BE107933 602278215
30	17.8	71.2	534	119	AA644558	AA644558 HS.5105_B
31	17.8	71.2	534	119	AA644558	AA644558 cm42e003.W
32	17.8	71.2	533	23	A1676750	A1676750 etmes7052
33	17.8	71.2	748	141	BE912977	BE912977 601668948
34	17.6	70.4	172	113	AM253374	AM253374 UI-R-BJ0-
35	17.6	70.4	194	111	AM121471	AM121471 UI-M-BH2-
36	17.6	70.4	244	103	A1905069	A1905069 QV-BT077-
37	17.6	70.4	255	222	FR0011711	AL002970 F.rubripde
38	17.6	70.4	310	134	BB460567	BB460567 BB460567
39	17.6	70.4	319	143	BF021048	BF021048 ux51f03.x
40	17.6	70.4	326	120	AM786182	AM786182 119087 MA
41	17.6	70.4	333	102	A1843362	A1843362 UI-M-AQ1-
42	17.6	70.4	336	139	BE766238	BE766238 IL3-NF010
43	17.6	70.4	345	5	AA317430	AA317430 EST19483
44	17.6	70.4	359	118	AM576427	AM576427 UI-HF-BP0
45	17.6	70.4	365	223	AQ045822	AQ045822 RPT111-35

ALIGNMENTS

RESULT 1
LOCUS A1514166 642 bp mRNA EST
DEFINITION GH27238.5prline GH Drosophila melanogaster head pot2 Drosophila
melanogaster cDNA clone GH27238 5prime, mRNA sequence.

ACCESSION A1514166
KEYWORDS A1514166.1 GI:4418228
SOURCE EST.
ORGANISM Drosophila melanogaster
fruit fly.

REFERENCE 1 (bases 1 to 642)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
P., Lewis, S., and Rubin, G. M.

AUTHORS BDGP/HMI Drosophila EST Project
TITLE

JOURNAL COMMENT

Unpublished (1997)
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
hit genomic sequence AC004313
Plate: 272 row: D column: 2
High quality sequence stop: 463.

FEATURES

source
1..642
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GH27238"
/clone_lib="GH Drosophila melanogaster head pot2"
/sex="male and female"
/dev_stage="adult"
/lab_host="DHS - alpha"
/note="Organ: head; Vector: pot2; Site:1: EcoRI; Site:2:
XhoI; Sized fractionated cDNAs were directly ligated into
pot2. Plasmid cDNA library."
BASE COUNT 147 a 182 c 183 g 1 others

Query Match 79.2%; Score 19.8; DB 21; Length 642;
Best Local Similarity 91.3%; Pred. No. 97;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 atctgctgttgagcgcagcag 25
|| ||||| ||||| ||||| |||||
DB 457 ATGTGCTGCTTGAAGCGCAGCAG 479

RESULT 2
LOCUS BE114310/c 507 bp mRNA EST
DEFINITION BE114310 13-JUN-2000
UI-R-BJ1-awk-g-02-0-UI-s1 UI-R-BJ1 Rattus norvegicus cDNA clone
UI-R-BJ1-awk-g-02-0-UI 3', mRNA sequence.
ACCESSION BE114310
VERSION BE114310.1 GI:8506415
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 507)
Bonaldo, M.F., Lennon, G., and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mssoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first A
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized heart library cDNA library Preparation: M.B. Soares Lab
Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=yes.
Location/Qualifiers
1..507

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-A0-bc-h-05-0-UI"
/clone_lib="UI-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; This library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dT track which allows identification of
the library of origin of a clone within the mixture."
BASE COUNT      125 a      119 c      105 g      158 t
ORIGIN

```

```

Query Match      77.6%; Score 19.4; DB 163; Length 507;
Best Local Similarity 95.2%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 4 tctgtctttgaagcgagca 24
|||||
Db 357 TCTGCTGTTGAAGCTCAGCA 337

```

```

RESULT 3
AA818607/c 528 bp mRNA EST 03-JUL-1999
LOCUS UI-R-A0-bc-h-05-0-UI.s1 UI-R-A0 Rattus norvegicus cDNA clone
DEFINITION Immunoglobulin-like receptor PIRAZ (6M23) mRNA, partial cds, mRNA
sequence.
ACCESSION AA818607
VERSION AA818607.1 GI:4228398
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 528)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
On Feb 17, 1998 this sequence version replaced gi:2889346.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult kidney library. cDNA Library Preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research
Genetics. This clone is also available through the I.M.A.G.E.
Consortium at LNL (info@image.lnl.gov). IMAGE ID=1768272
Seg primer: M13 forward
POLYA=No.
Location/Qualifiers
1..528
/organism="Rattus norvegicus"

```

```

/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-A0-bc-h-05-0-UI"
/clone_lib="UI-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; This library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dT track which allows identification of
the library of origin of a clone within the mixture."
BASE COUNT      138 a      131 c      101 g      158 t
ORIGIN

```

```

Query Match      77.6%; Score 19.4; DB 12; Length 528;
Best Local Similarity 95.2%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 4 tctgtctttgaagcgagca 24
|||||
Db 353 TCTGCTGTTGAAGCTCAGCA 333

```

```

RESULT 4
AO522448 437 bp DNA GSS 11-MAY-1999
LOCUS HS_5237_A2_A12_T7A RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=813 Col=24 Row=A, DNA sequence.
ACCESSION AO522448
VERSION AO522448.1 GI:4769482
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 437)
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 813 row: A column: 24
Seg primer: T7
Class: BAC ends
High quality sequence stop: 437.
Location/Qualifiers
1..437
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=813 Col=24 Row=A"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor

```

FEATURES

Source

```

FEATURES
Source
1..528
/organism="Rattus norvegicus"

```


RESULT 9
CNS02Y61/C
LOCUS
DEFINITION
CNS02Y61 802 bp DNA GSS 15-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
180C23 of library G from Tetraodon nigroviridis, genomic survey
sequence.
AL219267 GI:7878086
VERSION
KEYWORDS
SOURCE
ORGANISM
Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 802)
REFERENCE
AUTHORS
Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
TITLE
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL
REFERENCE
2 (bases 1 to 802)
AUTHORS
Roest-Crollius, H., Jallion, O., Dasilva, C., Bonneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLE
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL
REFERENCE
3 (bases 1 to 802)
AUTHORS
Genoscope.
TITLE
Direct Submission
COMMENT
Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
source
1..802
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99863"
/clone="180C23"
/clone_1lb="G"
/note="Genoscope sequence ID : COAG180AB12LPI-end : T7"
BASE COUNT
187 a 166 c 188 g 254 t 7 others
ORIGIN

Query Match 75.2%; Score 18.8; DB 220; Length 802;
Best Local Similarity 83.3%; Pred. No. 2.9e+02;
Matches 20; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 gatctgctgttgaagcgacag 25
1:1 ||||| ||||| ||||| |||||
Db 668 GWTGCTGCTGTTGCTGCGACAG 645

RESULT 10
AJ394746
LOCUS
DEFINITION
AJ394746 446 bp mRNA EST 25-JAN-2001
AJ394746 dktz426 Gallus gallus cDNA clone 119r1, mRNA sequence.
AJ394746
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
Chicken.
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 446)
REFERENCE
AUTHORS
Abdrakhmanov, I., Lodgyn, D., Gerolt, P., Atakawa, H., Law, A., Plachy

TITLE
JOURNAL
MEDLINE
COMMENT
',', Korn, B. and Buerstedde, J.M.
A large database of chickenursal ESTs as a resource for the
analysis of vertebrate gene function
Genome Res. 10 (12), 2062-2069 (2000)
20568495
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinstr. 52, 20251 Hamburg, Germany
Email: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.

FEATURES
source
1..446
/organism="Gallus gallus"
/strain="CB"
/db_xref="taxon:9031"
/clone="119r1"
/clone_1lb="dktz426"
/tissue-type="Bursa of Fabricius"
/dev-stage="tallbud"
BASE COUNT
122 a 107 c 117 g 100 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 104; Length 446;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggatctgctgttgaagcgacag 25
111 ||||| ||||| ||||| |||||
Db 407 GGAGCTGCTGTGGAAGTCAGAG 431

RESULT 11
AV680177
LOCUS
DEFINITION
AV680177 468 bp mRNA EST 05-OCT-2000
AV680177 Nori Satoh unpublished cDNA library Ciona intestinalis
cDNA clone rcitbln19 3', mRNA sequence.
AV680177
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
Ciona intestinalis.
Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Phlebobranchia; Clonidae; Ciona.
1 (bases 1 to 468)
REFERENCE
AUTHORS
Satoh, N., Satou, Y., Kohara, Y. and Shin, I.T.
TITLE
Expressed genes in Ciona intestinalis
JOURNAL
COMMENT
Unpublished (2000)
Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source
1..468
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="rcitbln19"
/clone_1lb="Nori Satoh unpublished cDNA library"
/tissue-type="whole animal"
/dev-stage="tallbud"
BASE COUNT
107 a 136 c 71 g 154 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 32; Length 468;
Best Local Similarity 84.0%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggatctgctgttgaagcgacag 25
111 ||||| ||||| ||||| |||||
Db 270 GGATCTGCTGTGGAAGTCAGAG 294

RESULT 12
LOCUS AV677254 509 bp mRNA EST 05-OCT-2000
DEFINITION AV677254 Nori Satoh unpublished cDNA library Ciona intestinalis
CDNA clone rcitb5g 3', mRNA sequence.
ACCESSION AV677254
VERSION AV677254.1 GI:10115253
KEYWORDS EST.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis.
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Phlebobranchia; Clonidae; Ciona.
REFERENCE 1 (bases 1 to 509)
AUTHORS Satoh, N., Satou, Y., Kohara, Y. and Shin-I, T.
TITLE Expressed genes in Ciona intestinalis
JOURNAL Unpublished (2000)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1..509
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone_lib="Nori Satoh unpublished cDNA library"
/tissue_type="whole animal"
/dev_stage="tailbud"

BASE COUNT 112 a 147 c 88 g 162 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 32; Length 509;
Best Local Similarity 84.0%; Pred. NO. 3.3e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgcagcag 25
|||||
Db 198 GGATCTGCTGTTGCAGCTAACAAG 222

RESULT 13
LOCUS AV675262 617 bp mRNA EST 05-OCT-2000
DEFINITION AV675262 Nori Satoh unpublished cDNA library Ciona intestinalis
CDNA clone citb1n19 5', mRNA sequence.
ACCESSION AV675262
VERSION AV675262.1 GI:10113261
KEYWORDS EST.
SOURCE Ciona intestinalis.
ORGANISM Ciona intestinalis.
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Phlebobranchia; Clonidae; Ciona.
REFERENCE 1 (bases 1 to 617)
AUTHORS Satoh, N., Satou, Y., Kohara, Y. and Shin-I, T.
TITLE Expressed genes in Ciona intestinalis
JOURNAL Unpublished (2000)
COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@ascidian.zool.kyoto-u.ac.jp.
Location/Qualifiers
1..617
/organism="Ciona intestinalis"
/db_xref="taxon:7719"

FEATURES
Source

BASE COUNT 187 a 130 c 169 g 130 t 1 others
ORIGIN

Query Match 74.4%; Score 18.6; DB 32; Length 617;
Best Local Similarity 84.0%; Pred. NO. 3.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgcagcag 25
|||||
Db 502 GGATCTGCTGTTGCAGCTAACAAG 478

RESULT 14
LOCUS CNS04A1A/c 904 bp DNA GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
095F11 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL281899.1 GI:8020229
VERSION AL281899
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 904)
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Bonneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 904)
AUTHORS Direct Submission
TITLE Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.
Location/Qualifiers
1..904
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="G"
/note="Genoscope sequence ID : CDBG095CC06SFI-end :
PUC-ori"

BASE COUNT 231 a 201 c 220 g 251 t 1 others
ORIGIN

Query Match 74.4%; Score 18.6; DB 221; Length 904;
Best Local Similarity 84.0%; Pred. NO. 3.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggaatcgtctgttgaagcgcagcag 25
|||||
Db 626 GGATCTGCTGTTGCAGCTCAACTG 602

```

RESULT 15
CNS03FNU 930 bp DNA GSS 17-MAY-2000
LOCUS Tetradon nigroviridis genome survey sequence pUC-ori end of clone
DEFINITION 02213 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL241912.1 GI:7962681
VERSION GSS: genome survey sequence.
KEYWORDS Tetradon nigroviridis.
SOURCE Tetradon nigroviridis.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 930)
Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 930)
Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
AUTHORS This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
1..930
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="02213"
/clone_lib="G"
/Note="Genoscope sequence ID : C0BG022CE07SP1-end :
pUC-ori"
LOCATION/Qualifiers

BASE COUNT 217 a 230 c 285 g 196 t 2 others
ORIGIN

Query Match 74.4%; Score 18.6; DB 221; Length 930;
Best Local Similarity 84.0%; Pred. No. 3.6e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgagcagc 25
|||||
Db 781 GGACCTGCTTTGAAGAAGAGCAGC 805

RESULT 16
CNS051BD/C 610 bp DNA GSS 26-JUL-2000
LOCUS Tetradon nigroviridis genome survey sequence 17 end of clone
DEFINITION 05101 of library A from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL316750.1 GI:9549634
VERSION GSS: genome survey sequence.
KEYWORDS Tetradon nigroviridis.
SOURCE Tetradon nigroviridis.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

```

```

Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 610)
Roest Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brotlier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Estimate of human gene number provided by genome-wide analysis
using Tetradon nigroviridis DNA sequence
JOURNAL Nat. genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
REFERENCE 2 (bases 1 to 610)
Roest Crolius,H.R., Jallion,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C.,
Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W.,
Bernot,A. and Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish tetraodon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
REFERENCE 3 (bases 1 to 610)
Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
AUTHORS This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
1..610
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"
/clone="05101"
/clone_lib="A"
/Note="Genoscope sequence ID : C0A051AH06C1-end : 17"
LOCATION/Qualifiers

BASE COUNT 124 a 159 c 162 g 139 t 26 others
ORIGIN

Query Match 73.6%; Score 18.4; DB 222; Length 610;
Best Local Similarity 95.0%; Pred. No. 4.2e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ggaatcgtctgttgaagcgac 20
|||||
Db 333 GGATCTCGCAGTTTGAAGCGC 314

RESULT 17
AV142481 185 bp mRNA EST 02-JUL-1999
LOCUS AV142481 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
DEFINITION clone 2810423K17, mRNA sequence.
ACCESSION AV142481
VERSION AV142481.1 GI:5346476
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 185)
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T., Hara
,A., Hayatsu,N., Horii,F., Ishikawa,T., Itoh,M., Izawa,M., Kawai,J.,
Kikuchi,N., Kojima,Y., Matsuyama,T., Niihuma,H., Oda,H., Owa,C.,
Sato,K., Shibata,Y., Shigenoto,Y., Shiraki,T., Sogabe,Y., Sugahara
,Y., Suzuki,H., Suzuki,H., Tateo,M., Tomaru,T., Tomioka,N.,
Watanabe,S., Yagame,M., Yamamura,T., Yokota,T., Yoshino,M.,
Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
RIKEN Mouse ESTs
JOURNAL Unpublished (1999)
CONTACT: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

```

Tel: 81-298-36-9145
Fax: 81-298-36-9098

Email: genome-res@rtc.riken.go.jp
Thermotabilization and thermostabilization of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES
source
Location/Qualifiers
1. 185
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="2810423K17"
/clone_lib="Mus musculus C57BL/6J 10-11 day embryo"
/sex="mixed"
/dev_stage="10-11 day embryo"

BASE COUNT
ORIGIN
51 a 46 c 38 g 50 t

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 25; Length 185;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 accgctgttgaagcgcagcag 25
||||| ||||| |||||
Db 41 ATCTGCTTTTGATCAGCAGCAG 63

RESULT 18
AA750666/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Oryza sativa.
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehretaceae; Oryzaceae; Oryza.
1 (bases 1 to 243)
Nahm, B.H., Kim, J.K., Cheong, J.J., Kim, S.I., Han, T.R., Moon, E.P.,
Kim, W.I., Kim, W.Y., Yang, M.S., Park, R.D., Sohn, U.I., Kang, K.Y., Lee,
M.C. and Eun, M.Y.
Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
Unpublished (1998)
Contact: Eun M.Y.
Department of Cytogenetics
National Inst. of Agri. Sci. and Tech, RDA
Suwon, Kyunggi-do, Korea
Tel: 82 331 290 0301
Fax: 82 331 290 0307
Email: myeun@sun20.asi.re.kr
Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji
University, Yongin, Korea, 449-728 bnamh@server.myongji.ac.kr
Seq primer: M13 Reverse Primer.
Location/Qualifiers
1. 243
/organism="Oryza sativa"
/cultivar="Milyang23"
/db_xref="taxon:4530"
/clone="IGS0248"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source

Location/Qualifiers
1. 243
/organism="Oryza sativa"
/cultivar="Milyang23"
/db_xref="taxon:4530"
/clone="IGS0248"
/clone_lib="Rice Immature Seed Lambda ZAPII cDNA Library"
/tissue_type="Immature Seed"
/dev_stage="5 days after pollination"
/lab_host="E. coli SOLR"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; Directional cDNA library inserted into lambda ZAPII

BASE COUNT
ORIGIN
vector at 5' end with EcoRI and 3' end with Xho I site."
65 a 60 c 62 g 56 t

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 11; Length 243;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 2 gactgctgttgaagcgcagcag 24
||||| ||||| |||||
Db 240 GATCTGCTGTGTGAGCTCACCA 218

RESULT 19
AV282092
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mus musculus.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 251)
Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F.,
Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai,
C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, T., Nakamura, M., Oda, H., Okazaki, Y.,
Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata,
Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,
Suzuki, H., Takahashi, F., Tateo, M., Tomioka, N., Tsunoda, Y.,
Watabiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T.,
Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Riken Mouse ESTs (Kono, H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp,
URL: <http://genome.rtc.riken.go.jp/>
Sasaki, N., Izawa, M., Watabiki, M., Ozawa, K., Tanaka, T., Yoneda, Y.,
Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki,
Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.
Location/Qualifiers
1. 251
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4933428B13"
/clone_lib="RIKEN full-length enriched, adult male testis
(DH10B)"
/sex="male"

FEATURES
source

Location/Qualifiers
1. 251
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4933428B13"
/clone_lib="RIKEN full-length enriched, adult male testis
(DH10B)"
/sex="male"

```

/tissue-type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Site_1: Salt, Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer 15'
GAGGAGAGAGATCCACAGCGCTTTTTTTTTTTTTTVN 3', cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence 15'
GAGGAGAGAGATTCGAGTAAATTAATTAATCCCCCCCCC 3'. cDNA
was cloned into the XhoI and BamHI sites. Vector: a
modified plasmid pUC(+) after bulk excision from lambda
FLC I. Cloning sites, 5' end: Salt, 3' end: BamHI."

```

Query Match	72.8%	Score 18.2	DB 28	Length 251
Best Local Similarity	87.0%	Pred. No. 4.6e+02		
Matches 20; conservative	0;	Mismatches 3;	Indels 0;	Gaps 0
QY	3	atctgctgtttgaagcgacgacg	25	
Db	104	ATCTGCTTGTGAATTCACGACG	126	

LOCUS	DEFINITION	EST	21-JUL-2000
BB452775/c	BB452775	268 bp	mRNA
BB452775	BB452775	RIKEN full-length enriched, 12 days embryo spinal ganglion	
	Mus musculus cDNA clone D130019M02.3	similar to X76772	M. musculus
	mRNA for ribosomal protein S3, mRNA sequence.		

ACCESSION	BB452775	GI:9348223
VERSION	BB452775.1	
KEYWORDS	EST	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE
AUTHORS
Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arahata, T., Carninci

P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawajiri, K., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Kiyosawa, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomioka, N., Toyota, T., Tanoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamanaka, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al.)

TITLE	RIKEN MOUSE ESTS (KONNO, H., ET AL.)
JOURNAL	UNPUBLISHED (2000)
COMMENT	CONTACT: YOSHIIHIDE HAYASHIZAKI

Genome Expression Research Group, Life Science Tsukuba Center,
Genome Science Laboratory
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-reserctc.riken.go.jp,
URL: <http://genome.ritc.riken.go.jp/>
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoke, S., Sasakawa,
N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Thermosensitization and thermoinactivation of thermolabile enzymes by

trehalose and its application for the synthesis of full length cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)

Itoch,M., Kitsuishi,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y. Automated filtration-based high-throughput plasmid preparation system. *Genome Res.* 9 (5), 465-470 (1999)

Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

Location/Qualifiers

1. 268

FEATURES	SOURCE
1. 268	Location/Qualifiers
	/organism="Mus musculus"
	/db_xref="taxon:10090"
	/clone="D130019M02"
	/clone_1lb="RIKEN full-length enriched, 12 days embryo spinal ganglion"
	/tissue_type="spinal ganglion"
	/dev_stage="12 days embryo"
	/lab_host="DH10B"
	/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGGAGAGAGCGCGCCGCACTCGAGTGTGTGTGTGTGTN 3'], cDNA was prepared by using triethanolamine thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGAGTTCGAGACTTAATTAATTAATTCGCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLIC I."
80 a	71 c 57 g 60 t
ORIGIN	BASE COUNT

Query Match	72.8%	Score 18.2	DB 134	Length 268
Best Local Similarity	87.0%	Pred. No. 4.7e+03		
Matches	20	Conservative	0	Mismatches 3
				Indels 0
				Gaps 0
QY	2	gacgcgcttttgaagcgacgca	24	
Db	28	GATCTGCACTTGAAGCACAGCA	6	

RESULT	21
LOCUS	AV216567/c
DEFINITION	AV216567 269 bp mRNA EST 30-OCT-1999 AV216567 RIKEN full-length enriched, ES cells Mus musculus cDNA clone 241016G18 3' similar to X51536 Rat mRNA for ribosomal protein S3, mRNA sequence.

ACCESSION	AV216567	GI:6157412
VERSION	AV216567.1	
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

[illegible]

Fukuda, S., Fukushima, Y., Harai, A., Hayashi, N., Hirozane, Y., Hori,
Shi, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Ka-
C. Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, A., Shibata, K., Shiga-
Y., Shigemoto, Y., Shiraki, T., Sogabe, I., Suganuma, T., Suzuki, H.,
Suzuki, H., Takahashi, F., Tateno, M., Tomimada, N., Tsunoda, Y.,
Tatani, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T.,

TITLE
JOURNAL
COMMENT

Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Konno, H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Genome Exploration Research Group, Life Science Tsukuba Center,
The Institute of Physical and Chemical Research (RIKEN), Genomic
Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: +81-298-36-9013
Fax: +81-298-36-9098
Email: genome-res@rtc.riken.go.jp
URL: http://genome.rtc.riken.go.jp/

Matsuda, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7): 3455-3460 (1998)
Itoh, M., Kikuchi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5): 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES
SOURCE

Location/Qualifiers
1..269

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1ib="2410164G18"
/clone_1ib="RIKEN full-length enriched, ES cells"
/cell_type="ES cells"
/lab_host="SOLR"
/note="Site 1: XhoI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGATCCAGACGCTTTTCTTTTCTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 5.0 and subtraction to Rot = 25.0. Second strand
cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGAGATTCGAGTTAATTAATATCCCCCCCCC 3']."

BASE COUNT
ORIGIN

79 a 71 c 60 g 59 t

Query Match

Best Local Similarity 72.8%; Score 18.2; DB 26; Length 269;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgagca 24
||||| 1 ||||| 1 ||||| 1
Db 39 GATCTTCAGTTTGATGCCGACCA 17

RESULT 22
AV136669/c 283 bp mRNA EST 01-JUL-1999
LOCUS AV136669 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
DEFINITION clone 2810027K19, mRNA sequence.
ACCESSION AV136669
VERSION AV136669.1 GI:5322399
KEYWORDS EST.
SOURCE house mouse.

ORGANISM

Mus musculus

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 283)
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara,
A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J.,
Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C.,
Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara,
Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tomioka, N.,
Matsuda, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,
Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermotabilization and thermostabilization of thermostable enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES
SOURCE

Location/Qualifiers
1..283

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1ib="2810027K19"
/clone_1ib="Mus musculus C57BL/6J 10-11 day embryo"
/sex="mixed"
/dev_stage="10-11 day embryo"

BASE COUNT
ORIGIN

75 a 74 c 69 g 65 t

Query Match

Best Local Similarity 72.8%; Score 18.2; DB 25; Length 283;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 gatctgctgttgaagcgagca 24
||||| 1 ||||| 1 ||||| 1
Db 54 GATCTTCAGTTTGATGCCGACCA 32

RESULT 23
LOCUS

AV214322/c 286 bp mRNA EST 30-OCT-1999
DEFINITION AV214322 RIKEN full-length enriched, ES cells Mus musculus cDNA
clone 2410135P16 3' similar to X51536 Rat mRNA for ribosomal
protein S3, mRNA sequence.
ACCESSION AV214322
VERSION AV214322.1 GI:6155168
EST.

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 288)
Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T.,
Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F.,
Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai,
C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M.,
Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,
Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata,
Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H.,
Suzuki, H., Takahashi, F., Tateo, M., Tomioka, N., Tsunoda, Y.,
Watanabe, S., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T.,

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
Location/Qualifiers

FEATURES

source

1. 481
/organism="Lotus japonicus"
/db_xref="taxon:34305"
/clone="WM013b10_1"
/dev_stage="young plants (two-week old)"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; isolate=Myakojima MG-20"

BASE COUNT
ORIGIN

104 a 113 c 115 g 149 t

Query Match 72.8%; Score 18.2; DB 30; Length 481;
Best Local Similarity 87.0%; Pred. No. 5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atctgtctgtttgaagcagcag 25
|||||
Db 240 ATCTGCTGTTGGAGCTCATCAG 218

RESULT 26
BG160914/c

LOCUS BG160914 584 bp mRNA EST 06-FEB-2001
DEFINITION des3c07.x1 Wellcome CRC PRN3 dorsal 1lp Xenopus laevis cDNA clone
IMAGE:3474037 3' similar to TR:Q15059 Q15059 KIAA0043 PROTEIN. ;
mRNA sequence.

ACCESSION BG160914 GI:12694833
VERSION BG160914.1
KEYWORDS EST.

SOURCE African clawed frog.
ORGANISM Xenopus laevis

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 584)
Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person
, B., Gibbons, M., Harvey, N., Rittler, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

TITLE
JOURNAL
COMMENT

Email: est@watson.wustl.edu
Library constructed by A.M. Zorn (Wellcome/CRC Institute). DNA
Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LNL at: info@image.llnl.gov
Possible reversed clone: similarity on wrong strand
Seq primer: -40UP from Gibco
High quality sequence stop: 494.

FEATURES
source

1. 584
/organism="Xenopus laevis"
/db_xref="taxon:8335"
/clone="IMAGE:3474037"
/clone_1lb="Wellcome CRC PRN3 dorsal 1lp"
/tissue_type="dorsal 1lp"
/lab_host="DH108 (phage-resistant)"
/note="Vector: pBSK3; Site_1: NotI; Site_2: EcoRI; CDNA
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library was constructed
by A.M. Zorn (Wellcome/CRC Institute)."
BASE COUNT 227 a 108 c 163 g 86 t
ORIGIN

Query Match 72.8%; Score 18.2; DB 174; Length 584;
Best Local Similarity 87.0%; Pred. No. 5.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcatctgtttgaagcagc 23
|||||
Db 470 GGATTGCTGTTTGGCGGACG 448

RESULT 27
A1325366 685 bp mRNA EST 23-DEC-1998
LOCUS A1325366/c
DEFINITION m129f01.y1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
clone IMAGE:464953 5' similar to TR:Q60947 Q60947 MAX-INTERACTING
TRANSCRIPTIONAL REPRESSOR. ; mRNA sequence.

ACCESSION A1325366 GI:4059795
VERSION A1325366.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 685)
Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The Washu-HMT Mouse EST Project
Unpublished (1996)
Contact: Maria M/Mouse EST Project
Washu-HMT Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

TITLE
JOURNAL
COMMENT

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
DOI: 278769
This read is a RESEQUENCE of a previously sequenced mouse clone.
This read has been verified (found to hit its original self in the
correct orientation)
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 384.

FEATURES
source

1. 685
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:464953"
/clone_1lb="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH108"
/note="Vector: p17T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
GGTTCACATCGAAGTGGAGCGGCGCGGCGAATTTTGTGTGTGTGTGTGTGT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 j; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p17T3 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Facina Bonaldo."
BASE COUNT 158 a 194 c 230 g 103 t
ORIGIN

Class: BAC ends
High quality sequence stop: 525.
Location/Qualifiers

FEATURES

source

1. 525

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Plate-681 Col-10 Row=D"

/clone_lib="RPC1-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

BASE COUNT 123 a 131 c 99 g 166 t 6 others

ORIGIN

Query Match 71.2%; Score 17.8; DB 229; Length 525;
Best Local Similarity 90.5%; Pred. No. 7.8e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tctgtgtttgaagcgcagca 24

Db 269 TGTGCTGTTGAAGCCAGCA 249

RESULT 31

AM644558

LOCUS AM644558 534 bp mRNA EST 03-APR-2000
DEFINITION cma2e03.w1 Blackshear/Soares normalized Xenopus egg library Xenopus

laevis cDNA clone PBX0141E03 5', mRNA sequence.

AM644558

AM644558.1 GI:7401949

EST.

ORGANISM African clawed frog.

Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;

Xenopodidae; Xenopus.

1 (bases 1 to 534)

Blackhear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G.

Jr., Moore, P.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman

J.W., Bonaldo, M.F., and Soares, M.B.

The NIHES Xenopus Maternal EST Project

Unpublished (2000)

Contact: Perry J. Blackshear

Office of Clinical Research and Laboratory of Signal Transduction

National Institute of Environmental Health Sciences

A2-05 NIHES, 101 Alexander Drive, Research Triangle Park, NC 27709,

USA

Tel: 919 541-4899

Fax: 919 541-4571

Email: black009@niehs.nih.gov

Clone is available through Research Genetics, Inc., 2130 Memorial

Parkway, Huntsville, AL 35901

phone 800-533-4363 ext.cdn, fax 256-536-9016 att.cdn, email

cdna@resgen.com

FEATURES

source

1. 534

/organism="Xenopus laevis"

/db_xref="taxon:8395"

/clone="PBX0141E03"

/clone_lib="Blackshear/Soares normalized Xenopus egg

library"

/sex="female"

/tissue_type="unfertilized egg"

/cell_type="unfertilized egg"

/dev_stage="unfertilized egg"

/lab_host="DH10B"

/note="Vector: p773-Pac; Site_1: EcoRI; Site_2: NotI;

PolyA-selected mRNA was prepared from unfertilized Xenopus

laevis eggs. The library was constructed in the vector

p773-Pac as described in Bonaldo, M.F., Lennon, G. and

Soares, M.B. 'Normalization and subtraction: two

approaches to facilitate gene discovery', Genome Research

6:791-806, 1996. The first strand synthesis used a

NotI-RT18 primer; double stranded cDNAs were ligated to

EcoRI adaptors, digested with NotI, and directionally

cloned into the NotI and EcoRI-digested p773-Pac vector.

The library contained approximately 7.2 x 10⁵

recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT 138 a 125 c 149 g 122 t

ORIGIN

Query Match 71.2%; Score 17.8; DB 119; Length 534;
Best Local Similarity 90.5%; Pred. No. 7.8e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 atctgtgtttgaagcgcagc 23

Db 283 ATCTCCTGTTGAAGTCACAC 303

RESULT 32

AI676750/c

LOCUS AI676750 553 bp mRNA EST 19-MAY-1999
DEFINITION etmEST0529 Eth1 Eimeria tenella cDNA clone etm1093 5', mRNA

sequence.

AI676750

AI676750.1 GI:4877230

EST.

ORGANISM Eimeria tenella.

Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;

Eimeria.

1 (bases 1 to 553)

Man, K.L., Chong, S.P., Ng, S.T., Tomley, F.M. and Jang, M.S.

Survey of genes in Eimeria tenella merozoite

Unpublished (1999)

Contact: Wan KL

Centre for Gene Analysis and Technology

Universiti Kebangsaan Malaysia

43600 UKM Bangi, Selangor DE, Malaysia

Tel: 6 03 8292997

Fax: 6 03 8293249

Email: kwan@pkirisc.cc.ukm.my

PCR Primers

FORWARD: T73

BACKWARD: T7

Seq primer: SK.

FEATURES

source

1. 553

/organism="Eimeria tenella"

/strain="Houghton"

/db_xref="taxon:5802"

/clone="etm1093"

/clone_lib="Eth1"

/dev_stage="second generation merozoite"

/lab_host="X11-Blue MRF"

/note="Vector: Lambda ZAP11; Site_1: EcoRI; Site_2: XhoI;

Second generation merozoites of E. tenella H were purified

by column chromatography and mRNA extracted using a

FastTrack kit (Invitrogen). cDNA was synthesized and a

uni-ZAP XR library was constructed using cDNA synthesis

kit, ZAP-cDNA synthesis kit and ZAP-cDNA GigaPack III

Gold cloning kit (Stratagene). The library was amplified

once through E. coli X11-Blue MRF."

BASE COUNT 133 a 143 c 164 g 112 t 1 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 23; Length 553;
Best Local Similarity 90.5%; Pred. No. 7.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ctgctgtttgaagcgcagcag 25
||||| ||||| |||||
Db 289 CTGCTGTTGAAGTCAGCAGC 269

RESULT 33
BE912977 748 bp mRNA EST 29-SEP-2000
LOCUS 601668948F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3968890 5',
DEFINITION mRNA sequence.
ACCESSION BE912977.1 GI:10410127
VERSION BE912977.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 748)
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbsr@mail.nih.gov

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9145 row: p column: 11
High quality sequence stop: 593.
Location/Qualifiers
1..748

FEATURES
source
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_1ib="NCI_CGAP_Mam1"
/tissue_type="tumor, biopsy sample"
/dev_stage="3 months, virgin"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT 147 a 150 c 201 g 250 t
ORIGIN

Query Match 71.2%; Score 17.8; DB 141; Length 748;
Best Local Similarity 90.5%; Pred. No. 8.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 ctgctgtttgaagcgcagcag 25
||||| ||||| |||||
Db 567 CTGCTGTTGAAGTCAGCAGC 587

RESULT 34
AM253374 172 bp mRNA EST 17-DEC-1999
LOCUS AM253374
DEFINITION UI-R-BJ0-aen-b-09-0-UI.s1 UI-R-BJ0 Rattus norvegicus cDNA clone
ACCESSION UI-R-BJ0-aen-b-09-0-UI 3', mRNA sequence.
AM253374

VERSION AM253374.1 GI:6596965
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 172)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mesares@blue.weeg.uiowa.edu

Oligo-dT track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seg primer: M13 Forward
POLY-A-NO.

FEATURES
source
Location/Qualifiers
1..172
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ0-aen-b-09-0-UI"
/clone_1ib="UI-R-BJ0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BJ0 library is a subtracted library derived from the UI-R-AAL, UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and UI-R-AG1 libraries. These libraries represent tissues from rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV canal at 15 dpc, and ventricle at 13 dpc. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.
TAG_LIB=UI-R-BJ0
TAG_TISSUE=ventricle at 15 dpc
TAG_SEQ=GTGTC"
BASE COUNT 30 a 51 c 38 g 53 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 113; Length 172;
Best Local Similarity 83.3%; Pred. No. 8.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 gatctgtgtttgaagcgcagcag 25
||||| ||||| |||||
Db 53 GATCTGCTGTTTCATGCCAGCAGC 76

RESULT 35
AM121471 194 bp mRNA EST 22-OCT-1999
LOCUS AM121471
DEFINITION UI-M-BH2.2-aon-g-01-0-UI.s1 NIH_BMAP_M.S3.2 Mus musculus cDNA clone
ACCESSION UI-M-BH2.2-aon-g-01-0-UI 3', mRNA sequence.
AM121471
VERSION AM121471.1 GI:6096804
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 194)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: westmail.nih.gov

0190-01 track not found, Not I site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares lab clone distribution: NIH BMAP cDNA clones will be made available by the means that is soon to be determined. When NIH determines the means for distribution of the BMAP cDNA clones, this record will be updated accordingly when that means is determined. The following repetitive elements were found in this cDNA sequence: 14-87, >GC-rich#low_complexity
 Seq primer: M13 Forward
 POLYA-No.

FEATURES

Source

1. 194
 Location/Qualifiers
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BH2.2-son-9-01-0-UI"
 /clone_lib="NIH_BMAP_M.S3.2"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site.1: Not I; Site.2: Eco RI; The NIH_BMAP_M.S3.2 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M.S3.2, NIH_BMAP_M.S2, NIH_BMAP_M.S1. The subtracted library (NIH_BMAP_M.S3.2) was constructed as follows: PCR amplified cDNA inserts from NIH_BMAP_M.S2 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the NIH_BMAP_M.S2 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxypatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH_BMAP_M.S3.2 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG_LIB=NIH_BMAP_M.S3.2
 TAG_TISSUE=CA3
 TAG_SEQ=CCACA"
 BASE COUNT 16 a 73 c 71 g 34 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 111; Length 194;
 Best Local Similarity 83.3%; Pred. No. 8.5e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 gcatcgtcgtttgaagcgagcaga 24
 ||||||||| | ||||| |
 DB 150 GGATCTGCTGCTGCGAAGCAGCAGCA 127

RESULT 36
 LOCUS A1905069
 DEFINITION QV-BT077-120299-159 BT077 Homo sapiens cDNA, mRNA sequence.
 ACCESSION A1905069
 VERSION A1905069.1 GI:6495456
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 244)
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zaglo,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/seq/gethtml.pl?l=QVat2-QV-BT077-159.html&t3=120299&t4=1>)
 Seq primer: puc 18 forward.

FEATURES

Source

1. 244
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT077"
 /sex="female"
 /dev_stage="Adult"
 /note="organ: breast; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 39 a 70 c 85 g 50 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 103; Length 244;
 Best Local Similarity 83.3%; Pred. No. 8.7e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 gatcgtcgtttgaagcgagcag 25
 ||||||||| | ||||| |
 DB 65 GATCTGCTGCGCAAGAGCAGCAG 88
 RESULT 37
 FR0011711
 LOCUS FR0011711 255 bp DNA
 DEFINITION F.rubripes GSS sequence, clone 065H13C12, genomic survey sequence.
 ACCESSION AL002970
 VERSION AL002970.1 GI:2448540
 KEYWORDS GSS: genome survey sequence.
 SOURCE Takifugu rubripes.
 ORGANISM Takifugu rubripes
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;

REFERENCE	TITLE	JOURNAL	COMMENT
1 (bases 1 to 255)	Acinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.	Submitted (09-SEP-1997) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. Email: biohelp@hmp.mrc.ac.uk	Vector: pBluescript II KS V type: phagemid PRIMER: KS
2	gacgcgtgttggaagcagcagc 25		
38	BB460567/c		
LOCUS	BB460567		
DEFINITION	BB460567 RIKEN full-length enriched, 12 days embryo spinal ganglion mus musculus cDNA clone D130067102.3' similar to X76772 M. musculus mRNA for ribosomal protein S3, mRNA sequence.		
ACCESSION	BB460567		
VERSION	BB460567.1		
KEYWORDS	GI:9356060		
SOURCE	EST.		
ORGANISM	mus. mouse.		
REFERENCE	Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus. 1 (bases 1 to 310)		
AUTHORS	Konno, H., Aizawa, K., Akahira, S., Akiyama, T., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, D., Ishikawa, T., Itoh, M., Iizawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomiura, N., Toya, T., Tsunoda, Y., Watabiki, A., Watanabe, S., Yamamura, T., Yamana, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.		
TITLE	RIKEN Mouse ESTs (Konno, H., et al.)		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Yoshihide Hayashizaki Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan Tel.: +81-298-36-9013 Fax: +81-298-36-9098 Email: genome-res@tc.riken.go.jp, URL: http://genome.rtc.riken.go.jp/ Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,		

N. Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.,
Thermolabile enzymes by
tRNAse and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kikunishi, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,
Y. and Hayashizaki, Y.,
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.,
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.

FEATURES
SOURCE

1. 310
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="D130067102"
/clone_lib="RIKEN full-length enriched, 12 days embryo
spinal ganglion"
/tissue_type="spinal ganglion"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Science Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGCGCCGACGATTTTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalase thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCGAGTTTAAATTAATTCGCCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
plasmidscript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT

88 a 91 c 62 g 69 t

ORIGIN

Query Match 70.4%; Score 17.6; DB 134; Length 310;
Best Local Similarity 83.3%; Pred. No. 9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0

QY 2 gatctgctgtttgaagcgacgacg 25
||||| ||||||| |||||||
Db 77 GATCTTCACTTTGAGCCCGACGAG 54

RESULT 39
BF021048/c

LOCUS

DEFINITION

BF021048 319 bp mRNA EST 10-OCT-2000
uv51f03.x1 Soares NKMD mandible Mus musculus cDNA clone
IMAGE:3513821 3' similar to TR-092115 Q92115 CALCIUM CHANNEL
ALPHA-2-DELTA-C SUBUNIT. ;, mRNA sequence.

ACCESSION

BF021048

VERSION

BF021048.1 GI:10752380

KEYWORDS

EST.

SOURCE

house mouse.
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 319)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-femail.nih.gov
This clone is available royalty-free through LNL ; contact the

JOURNAL

COMMENT

BASE COUNT 63 a 101 c 107 g 62 t
ORIGIN TAG-SEQ-CAGAC"

Query Match 70.4%; Score 17.6; DB 102; Length 333;
Best Local Similarity 83.3%; Pred. No. 9e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggaatcgtctgtgaagcagca 24
150 GGATGCTGCTGTGAGCGACCA 127

RESULT 42
BE766238 336 bp mRNA EST 19-SEP-2000
LOCUS IL3-NT0104-270600-187-B07 NT0104 Homo sapiens cDNA, mRNA sequence.
DEFINITION BE766238
ACCESSION BE766238.1 GI:10196162
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 336)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

Shoqun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
COMMENT 20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3-NT0104-270
600-187-B07&tl3=2000-06-27&tl=1)
Seq primer: puc 18 forward
High quality sequence stop: 336.

FEATURES

source Location/Qualifiers
1..336
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0104"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: Sma1;
Site_2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 68 a 95 c 66 g 107 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 139; Length 336;
Best Local Similarity 83.3%; Pred. No. 9.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggaatcgtctgtgaagcagca 24
||||||| ||||| |||||

Db 185 GGATGCTGCTGTGAGCGACCA 208

RESULT 43
AA317430 345 bp mRNA EST 19-APR-1997
LOCUS EST19483 Retina II Homo sapiens cDNA 5' end similar to EST
DEFINITION containing Alu repeat, mRNA sequence.
ACCESSION AA317430
VERSION AA317430.1 GI:1969871
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 345)
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult
C.J., Lee, N.H., Kirtness, E.F., Weissstock, K.G., Gocayne, J.D., White
O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Val, C., Clayton, R.A.,
Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald
L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S., Glodek, A.,
Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M.,
Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palacios, R.E., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
Dincke, D., Feng, D.-F., Ferrite, A., Fischer, C., Hastings, G.A., He, W.W.,
Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L.,
Kunsch, C., Hungjun, J., Li, H., Melssner, P.S., Olsen, H., Raymond, L.,
Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon
M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
Venter, J.C.

Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 suppl), 3-174 (1995)
COMMENT 96026280
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tadb/hgi/hgi.html)
Seq primer: M13 Reverse.

FEATURES
source Location/Qualifiers
1..345
/organism="Homo sapiens"
/db_xref="ATCC (lnst):118001"
/db_xref="taxon:9606"
/clone_lib="Retina II"
/sex="mixed"
/dev_stage="adult"
/note="Organ: retina; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"

BASE COUNT 96 a 80 c 81 g 87 t 1 others
ORIGIN

Query Match 70.4%; Score 17.6; DB 5; Length 345;
Best Local Similarity 83.3%; Pred. No. 9.1e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggaatcgtctgtgaagcagca 24
||||||| ||||| |||||

Db 67 GGAGTGTCTGTATGTAGCAGCA 90

RESULT 44
AM576427/c

LOCUS AM576427 359 bp mRNA EST 15-MAR-2000
 DEFINITION UI-HE-BP0P-aja-d-10-0-UI.s1 NIH_MGC_51 Homo sapiens cDNA clone
 IMAGE:3073602 3', mRNA sequence.
 ACCESSION AM576427
 VERSION AM576427.1 GI:7247966
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 359)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Oligo-4T track not found. Not 1 site shown in beginning of sequence
 is likely internal to the message. Tissue Procurement: Louis M.
 Staudt, M.D., Ph.D.
 cDNA Library Preparation: M.B. Soares Lab
 cDNA Library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNL at:
www.bio.lnl.gov/bbrp/image/image.html
 The following repetitive elements were found in this cDNA
 sequence:
 195-270, >GC-rich#low_complexity
 Seq primer: M13 Forward
 POLYA-No.

FEATURES
 source Location/Qualifiers
 1..359
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3073602"
 /clone_id="NIH_MGC_51"
 /tissue_type="lymph"
 /cell_type="germinal center B cells"
 /cell_line="MGC85"
 /lab_host="DH10B (LTR)"
 /note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
 Constructed from size fractionated cytoplasmic mRNA
 (4.4-7.4kb). Directionally cloned. Cells provided by
 Louis M. Staudt, Ph.D. Library preparation by Maria de
 Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 32 a 137 c 133 g 57 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 118; Length 359;
 Best Local Similarity 83.3%; Pred. No. 9.1e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 gatctgctgtttgaagcgacag 25
 ||||| | ||||| |||||
 Db 38 GACCTGCCGCTGAAGCGACGAG 15

RESULT 45
 A0045822/c
 LOCUS
 DEFINITION A0045822 365 bp DNA GSS 14-APR-1999
 R0111-35J21.TJ R011-11 Homo sapiens genomic clone R011-35J21,
 DNA sequence.
 ACCESSION A0045822
 VERSION A0045822.1 GI:3314749
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 365)
 Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
 Golden,K., Bery,K., Granger,D., Suh,E., Wible,C., de Jong,P. and

TITLE Venter, J.C.
 JOURNAL Use of BAC End Sequences for Sequence-Ready Map Building (1998)
 COMMENT Unpublished (1998)
 Other_GSSs: R0111-35J21.TK
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are derived from the human BAC library R011-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@tigr.org, med.buffalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from
 Research Genetics (inforesgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
 Class: BAC ends.

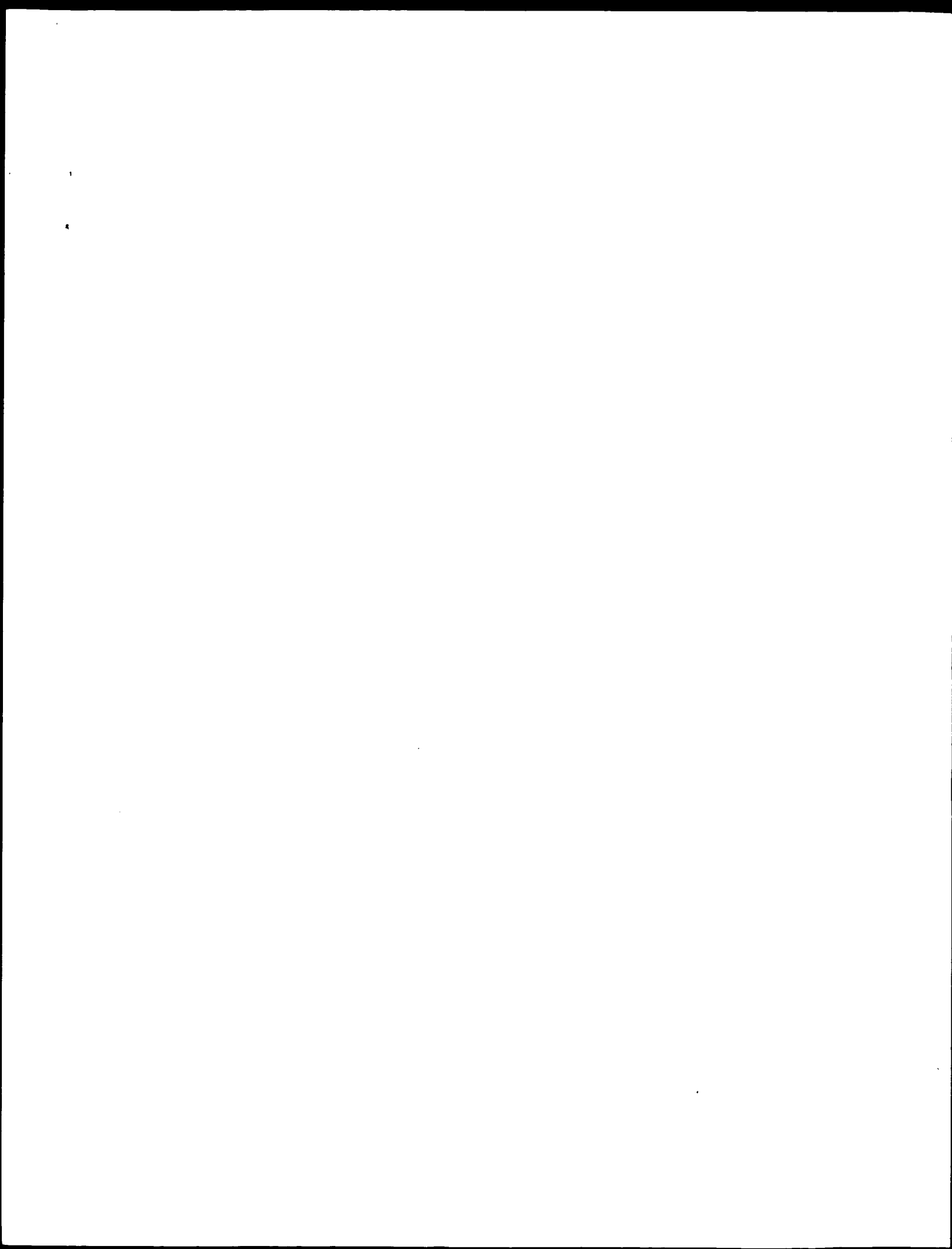
FEATURES
 source Location/Qualifiers
 1..365
 /organism="Homo sapiens"
 /db_xref="GDB:7513292"
 /db_xref="taxon:9606"
 /clone_image="R011-35J21"
 /clone_id="R011-35J21"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
 R0111 Human Male BAC Library"

BASE COUNT 79 a 79 c 85 g 122 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 223; Length 365;
 Best Local Similarity 83.3%; Pred. No. 9.1e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggatctgctgtttgaagcgacga 24
 ||||| | ||||| |||||
 Db 256 GGAGCTTCTGTGGGAAGCGACGCA 233

Search completed: October 9, 2001, 13:46:36
 Job time: 9511 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 12:11:18 ; Search time 1666.31 Seconds
(without alignments)
232.066 Million cell updates/sec

Title: us-09-396-196f-5

Sequence: 1 gcgcagcaggtgcatgcgcagcalt 25

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_in4:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_fod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_fod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_ov:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_r01:*
95: gb_r02:*
96: gb_r04:*
97: gb_pr10:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	9	AR029499
2	25	100.0	1041	9	AR034916
3	25	100.0	1084	9	A11530
4	25	100.0	5793	2	ECOBIO
5	25	100.0	5872	9	A38246
6	25	100.0	5872	9	A38251
7	25	100.0	5872	9	A93674
8	25	100.0	5872	9	A93679

9 25 100.0 5872 9 AR101809
 10 25 100.0 5872 9 AR101810
 11 25 100.0 11022 1 AR100180
 12 25 100.0 13501 1 AE000528
 13 25 100.0 297816 2 AP002553
 14 23.4 93.6 1121 10 E00893
 15 23.4 93.6 5526 2 AE250776
 16 20.4 81.6 12211 1 AE004895
 17 20.2 80.8 123926 14 R10081
 18 20.2 80.8 123926 76 AC079633
 19 20 80.0 2923 3 MAV250020
 20 19.2 76.8 12891 1 AE004192
 21 19.2 76.8 79419 93 HSJ324017
 22 19.2 76.8 349116 2 AP003003
 23 19 76.0 44201 3 MRCY180
 24 19 76.0 44201 3 PSEHPR1
 25 18.8 75.2 45704 1 AE001827
 26 18.8 75.2 110000 80 AL359456-2
 27 18.8 75.2 148900 69 AC025404
 28 18.8 75.2 176687 80 AL365213
 29 18.6 74.4 1359 13 AF250770
 30 18.6 74.4 1359 13 AF289025
 31 18.6 74.4 3162 15 SLDPHAI
 32 18.6 74.4 3175 12 AF029256
 33 18.6 74.4 3224 15 TOBPM3A
 34 18.6 74.4 3229 15 TOBPM3A
 35 18.6 74.4 3233 15 TOBPM3B
 36 18.6 74.4 3244 13 AF275745
 37 18.6 74.4 3426 14 MCB4891
 38 18.6 74.4 10302 13 AF179442
 39 18.6 74.4 33204 65 AC017995
 40 18.6 74.4 88548 13 AT27H5
 41 18.6 74.4 103787 14 ATTB10
 42 18.6 74.4 108879 13 AT12K4
 43 18.6 74.4 120181 63 AC014784
 44 18.6 74.4 184278 4 AC008221
 45 18.6 74.4 186549 4 AC008220

ALIGNMENTS

29-SEP-1999

RESULT 1
 AR029499 1041 bp DNA
 LOCUS AR029499
 DEFINITION Sequence 7 from patent US 5859335.
 ACCESSION AR029499
 VERSION AR029499.1 GI:5941472
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1041)
 AUTHORS Patton,D.Andrew.
 TITLE Enhanced biotin biosynthesis in plant tissue
 JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
 FEATURES
 source Location/Qualifiers
 1..1041
 /organism="unknown"
 BASE COUNT 262 a 273 c 305 g 201 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gccgacaggtgcatgcgcagcatt 25
 ||||||||||||||||||||
 Db 79 gccgacaggtgcatgcgcagcatt 103

RESULT 2

AR034916 AR034916 1041 bp DNA PAT 29-SEP-1999
 LOCUS AR034916
 DEFINITION Sequence 7 from patent US 5869719.
 ACCESSION AR034916
 VERSION AR034916.1 GI:5950521
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1041)
 AUTHORS Patton,D.A.
 TITLE Transgenic plants having increased biotin content
 JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
 FEATURES
 source Location/Qualifiers
 1..1041
 /organism="unknown"
 BASE COUNT 262 a 273 c 305 g 201 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gccgacaggtgcatgcgcagcatt 25
 ||||||||||||||||||||
 Db 79 gccgacaggtgcatgcgcagcatt 103

RESULT 3
 AR11530 AR11530 1084 bp DNA PAT 09-FEB-1994
 LOCUS AR11530
 DEFINITION BioB gene of E.coli with primers.
 ACCESSION AR11530
 VERSION AR11530.1 GI:490218
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 1084)
 AUTHORS
 JOURNAL Patent: GB 2216530-A 16 11-OCT-1989;
 FEATURES Location/Qualifiers
 source 1..1084
 /organism="Escherichia coli"

gene /db_xref="taxon:562"
 24..1064
 /gene="BioB"
 24..1064
 /gene="BioB"
 /codon_start=1
 /transl_table=11
 /protein_id="CAA00965.1"
 /db_xref="GI:490219"

/db_xref="SWISS-PROT:P12996"
 /translation="MAHPRMTLSVTFLEFKPLLDLLEFAQVYHRQHPDPROVYST
 LLSITGACPEDKTCPOSSRYKTGLEAEIRIMEVYESAKRAKAAGSTRFCGAAW
 KNPERMDPVLEQWQVGRKAMGLACMTLGLSSQORLANAGLDVYNNHLDPSEF
 YGNITTRTYOERLDLTLEKVDAGIKVCSGIVGLGVKDRAGILDLQANLPPPS
 VPINMLVVKGTPLADNDVDVAFEFIRIAVARIMPTSVYRSAGREONNEOTQAMC
 FMAGNSIFYGCKLITTPNPEKDLQLFRLTGLNPGQTVNLADNQCQRLEBALMT
 PTDVEYMAAL"

BASE COUNT 271 a 286 c 318 g 209 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gccgacaggtgcatgcgcagcatt 25
 ||||||||||||||||||||

Db 102 GCGCAGCAGTGCATCGCAGCATT 126

RESULT 4
LOCUS ECOBIO
DEFINITION E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc protein, and dehydrobiotin synthetase (bioD), complete cds.
ACCESSION J04423
VERSION J04423.1 GI:145422
KEYWORDS 7,8-diamino-pelargonic acid aminotransferase; 7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene; bioc gene; biocD gene; bioF gene; biotin synthetase; dehydrobiotin synthetase.
SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 5793)
AUTHORS Otsuka,A.J., Buonocristiani,M.R., Howard,P.K., Flamm,J. and Johnson,O.
TITLE The Escherichia coli biotin biosynthetic enzyme sequences predicted
J. Biol. Chem. 263, 19577-19585 (1988)
MEDLINE 89066784
COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.
FEATURES
source location/Qualifiers
1..5793
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
complement(98..574)
/note="ORF 1"
/codon_start=1
/transl_table=11
/protein_id="AA023513.1"
/db_xref="GI:145168"
/translation="MKLISNDLRDGLPHRHVFNMGQYDGNISPHLAMDVPAGTK SPVVCYDPAADATGSGMMHVVYVNLPAIDRVLPQCGFSGVAMPDGVLOTRTDFGKGTG YGAAPPKGETRHRYITFYHALDIERLDVDEGASAAVGVNHFHSLASASTAMFS"
complement(633..1925)
/gene="bioA"
/gene="bioc"
complement(633..1925)
/gene="biob"
/gene="biob"
/codon_start=1
/transl_table=11
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AA023514.1"
/db_xref="GI:145106"
/translation="MTTDLADPDRIHIMPTYSMTSPILVYVNVSAEGCELLISDGRRLVDCSSMAAIIHGVNHPOLINAAKMSQIDAMSHVNGGITHAPALELCKILVAMSGRN ALECVLADSGVAIVNHPOLINAAKMSQIDAMSHVNGGITHAPALELCKILVAMSGRN SMHSLMKGYLPENLAFAPAROSMDGEMDRDVGAFRAAARHETIAAYITIPDIYGA GGRMYHPEWIKRIRKICDRBGLIADIDINLFGFTGKFLACENHETIADIDICGKA LTGGTMTLSATLTREVAFITSGEAGCTMHGTFGNGFLACAAANASALILIESGMO QOYADIEVOLREQLAPARDAENAVADVIRGAVETTPVNNALOKFVBOGWIR PFGLITLMPYIILPQOLRLTAAVNAVODETFPCQ"
2012..3052
/gene="biob"
2012..3052
/gene="biob"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AA023515.1"
/db_xref="GI:145425"
/translation="MAHRRPWTLSQVTELEFKPLDLLEAOQVHROHDPQVQYST LLSIKTGACPEDCKYCPQTSRYKTYGLAEKLEMEVEVLSARAKAGSTREMGAM KNPHERMPYLEQMVGVKAMGLEACMTGLTISQOARLANAGVYNNHNDLTPSEF YGNITTRVYOEKLDITLEKVRDAGIKVCSGIVGLGETVYKDRAGLLIOLANLPTPES VPINMLVKVKGTPLADNDVDADPFIIRITIAVARIMPTSVIRLSAGREQMDQTOAMC

gene
CDS
3049..4203
/gene="biob"
3049..4203
/gene="biob"
/codon_start=1
/transl_table=11
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AA023516.1"
/db_xref="GI:145426"
/translation="MSWQEKINADLARRADALRRRYPAQAGHMLVADDROYLNF SSNDYGLSHHPQITRAMOQAGOFQSGSGSVAGSVYHOALEEAEMLVYSRA LITFSGCAANQAVIYAMMAKEDRIADRLSHASLLEASLSPSLRFPANVDYTHAR LLSAPCGQOMVYTEGYSMDGSAPIAEIQOVTQOCHGAMVDADHAGCTVYGEQGRG SCWLOKVPPELVTEGKFGVSGAVALCEQVADVLDLPARHLITSTSPRQOAL RASLAVIRSDGCAAREKLAULTIRFAGQDLPFLASCSA10PLVGDNSRALQ LAERLOQGWTA1RPTVPAGTARLTLTAHEMODIDRLLEVLHGNG"
4190..4945
/gene="bioc"
4190..4945
/gene="bioc"
/codon_start=1
/transl_table=11
/protein_id="AA023517.1"
/db_xref="GI:145427"
/translation="MATVKNQATAAFGRAAHYEOHADROSADALLAMPORYT HVIDAGGCGPMGMSRHRERAOYATDLSPMLVOAKQKRAHHTYLAGDIESLPLATA TFDLANSNLAVOWCGNLSTALRELRYVRKGVAAFTTLVQSLPERHQAQVDSER HANRELPDEIEOSLNGVYQHNIQPTLTFMEDALSMSRLKIGATHLHEGRDPRIL TRSOLRLOLAMPQOGRVPLTYHLFGVIARE"
4938..5597
/gene="biob"
4938..5597
/gene="biob"
/codon_start=1
/transl_table=11
/protein_id="AA023518.1"
/db_xref="GI:145428"
/translation="MSKRYEVTGDTVEGKTVASCALLQAKRAGTYTAKYKVASG EKTPREGLSNALALORNSSLOLDYAVNPEYFAEPSPHITSQDEGRPTESLVMSAG LRALEQOQADWVIVGAGGMPPLSDTFTFEDMWYQEDLPVILVGVVAGCINAMTLA QVIOHAGLTLAGVANDVTPPKRHAEMTTLTBMIPRCWERSPLQKIQKMRPES T"
BASE COUNT 1363 a 1554 c 1631 g 1245 t
ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match
Best local Similarity 100.0%; Score 25; DB 2; Length 5793;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgacagtgatgcgcagcatt 25
|||||
Db 2090 GCGCAGCAGTGCATCGCAGCATT 2114
|||||

RESULT 5
LOCUS A38246
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION A38246
VERSION A38246.1 GI:2294844
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,U., Fuhmann,M. and Shaw,N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 1 14-APR-1994;

COMMENT LONZA AG (CH)
 other publication PL 308301 950724
 other publication CA 2145400 940414
 other publication AU 4820293 940426
 other publication HU 71781 960228
 other publication SK 42095 951108
 other publication CZ 9500809 950913
 other publication FI 951547 950331
 other publication JP 8501694T 960227.
 location/Qualifiers

FEATURES
 source
 1..5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PBO30A-15/9"
 1..96
 /function="PROMOTER PTAC"
 23..28
 /evidence=experimental
 45..49
 /standard_name="PROMOTER PTAC"
 /evidence=experimental
 105..119
 /standard_name="BIOB RBS NO.9"
 /evidence=experimental
 117..1157
 /gene="BIOB"
 117..1157
 /gene="BIOB"
 /codon_start=1
 /transl_table=11
 /number=1
 /evidence=experimental
 /product="BIOTIN SYNTHASE"
 /protein_id="CAO2325.1"
 /db_xref="GI:2294845"
 /translation="MAHRPWTLSQVTELEKPLDLLFEAOQVHRHPPROVOVST
 LLSITKGCAPEDCKYCPOSSRYKTGLEAERLMEVOYLESARKKAGSTRPCGAAM
 KNPHRDMPTLEOVGVKAMGLACMTLGTSOQRLANGLDVYNNLDPSPF
 YGNIITRTYQERLDLLEKVDAGIKVCSGIVGCTVDRAGLLIQLANLPTPES
 VPIMLVKVGSTPLADNDVADPFRITIAVARIMFTSVYRISAGRENEQOQAC
 EMACANSIFPGCKLITTPNEPEDKDLQFRKLGILNPOOTAVLAGDNEQOQRLQALMT
 PDIDEYTNAAAL"
 2284..2297
 /standard_name="BIOC RBS"
 2295..3050
 /gene="BIOC"
 2295..3050
 /gene="BIOC"
 /gene="BIOC"
 /function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=3
 /product="PROTEIN"
 /protein_id="CAO2326.1"
 /db_xref="GI:2294846"
 /translation="MATVNRQALIAAFGRAAHYEDHADLROSADALLAMLPORRYT
 HVDAGCGPGWMSRHRERHAQVYALDLSRPMVVOAROKDAADHYLAGDIESLPLATA
 TFDLAMSNIWOMGNGNLSTALRELRYVRKGVAVATLIVOGSLPELHOMQVADVERP
 HANFLPDEIEOSINGVYOHIIOPITTFEDDALSMRSLKIGATHLHESGDRPIL
 TRSOLQRLAWAQOQGRYPLVYHFLGVIAFR"
 3742..3752
 /standard_name="BIOA RBS"
 3750..3039
 /gene="BIOA"
 3750..3039
 /gene="BIOA"
 /standard_name="S-ADENOSYL-L-METHIONINE:
 8-AMINO-7-OXONONANOATE AMINOTRANSF."
 /EC_number="2.6.1.62"
 /codon_start=1
 /transl_table=11
 /number=5

/evidence=experimental
 /product="DAPA SYNTHASE"
 /protein_id="CAO2327.1"
 /db_xref="GI:2294847"
 /translation="MTTDDIAFPQRIHMPYTSWISPLPYVVSAGECELLSDGRR
 LVQMSWMAAITHGVNHPOLNAAMKSQIDAMSHVMGGITTHAPALELCKRLVAMPPOP
 LECVFLADSGSVAVEMAKMALQYQAKGABRQRLTFPFGYHGDIFGAMSYCDPDS
 MSLMKGYLPEENLEFAPQSRMDGEMDERDVGAFARLMAHREHIAVATIEPIYOGAG
 GMRVHPEWTKRIRKICDREGILLIADETATGFGRTGLPACHEHAIAPDIICIGAL
 TGGTMTLSATLTITREVAETISNGEACGFMHGPTEKNGNPLCAAAANSLATILFSGDMOQ
 OVADIEVQAREQLPARDAEMVAVYLAIGVETTHRYVNAALQKFEVEGVNTRP
 EKLITLPEPYITLPOOLQRLTAAVNRVAVODETFEQ"
 5088..5100
 /standard_name="ORF1 RBS"
 5098..5574
 /gene="ORF1"
 5098..5574
 /gene="ORF1"
 /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=6
 /evidence=experimental
 /product="PROTEIN"
 /protein_id="CAO2328.1"
 /db_xref="GI:2294848"
 /translation="MKLISNDLRDGDKLPHRHVFNQMGYDGNISPHLAMDVPAGTK
 SFVYTCYDDAPATGSGMHHVYVNLPAFDRVLPQGFSGSLVAMPQGVLOTTRDFKGTG
 YDGAAPPKGETHRYIPTVHALDIERIDVDEGASGAMGVNHFHSLASATAMS"
 5583..5605
 5583..5644
 /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
 TERMINATOR"
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgacgagtgatcgccagcatt 25
 |||
 Db 195 GCGCAGAGGTGCATCGCCAGCATT 219

RESULT 6
 LOCUS A38251 5872 bp DNA PAT 05-MAR-1997
 DEFINITION Sequence 6 from Patent WO9408023.
 ACCESSION A38251
 VERSION A38251.1 GI:2294849
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch O., Brass J., Fuhrmann M. and Shaw N.
 TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 JOURNAL Patent: WO 9408023-A 6 14-APR-1994;

COMMENT LONZA AG (CH)
 other publication PL 308301 950724
 other publication CA 2145400 940414
 other publication AU 4820293 940426
 other publication HU 71781 960228
 other publication SK 42095 951108
 other publication CZ 9500809 950913
 other publication FI 951547 950331
 other publication JP 8501694T 960227.
 location/Qualifiers
 1..5872
 /organism="Escherichia coli"


```

/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
CDS
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/transl_table="MSMOEKNALADRADALRRRYVPAOGAGRWLVADROYLNF
SENDYGLSHRPQITRAMQOGAEOFGISGSGHYSGYSYHQALEELERLGYRA
LFIISFANQAVIAMAARKEEDRIADRLSHASLEASLSFSQLRRAHNDVTHLAR
LLASPCGQOMVVEGVFSMDGSDASPLAEIQVTOHNGMLVDDAHGTGVIGEGRG
SCMLQKVKPELLVTFEGKGVGAVALICSSVADYLLQFAHLIYSTMPAQAOL
RASLAVIRSEDEGADAREKLAALITFRAGVODLPPTLADSCSAIQPLIYCDMSRALQ
AEKLROCGCWTAIRPTYPATGATRLRLITNAHMQDIDRLLEVLHGNG"
3030..3045
RBS
/standard_name="BIOD RBS"
3043..3753
/gene="BIOD"
3043..3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DPR SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/transl_table="MSKRYFTVGTDTVEGKTVASCALLQAAAGRTAGYKPVASGS
EKTEPEGLRNSDALQORNSLQIDYATVNPYFAEPTSPHISAOEGRIEISLVASG
LRLAEQADMLVVEGAGGEMPTLSDFTFADWVTOQLPVLLVGVKILGICINHAMLTA
QVIOHNGCLTAGVANDVYPPGKRHAHEVMTITRLMIPALLGEIPWLAENPENATKG
YNLAFAVDSSTIGFTSRL"
1318 a 1352 c 1695 g 1307 t
BASE COUNT
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gcgcagcaggtgcacgcagcatt 25
111111111111111111111111
Db 195 GCCGACGAGGTGCATCCGCACGATT 219

```

```

RESULT 7
LOCUS A93674 5872 bp DNA PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS
SOURCE
Escherichia coli.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (bases 1 to 5872)
AUTHORS Birch,O. and Brass,J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 1 01-Oct-1997;

```

```

FEATURES
LOCATION AG (CH)
SOURCE
1..5872 Location/Qualifiers
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"
1..96
/promoter="PROMOTER PTAC"
/evidence=experimental
23..28
-35_signal
-10_signal
RBS
/standard_name="PROMOTER PTAC"
105..119
/evidence=experimental
/standard_name="BIOD RBS NO.9"
117..1157
/evidence=experimental
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/transl_table="MAHRPRWTLISQVTEIFKEPLDILFPAQOVHROHPPOVOYT
LLSIRKTCAPEDCKYCPPOSSRYKTGLAEERLMEYEOYLEARRAKKAGSRPFRCGAM
KNPIERDMPYLEOVGVKAMGLEACMTLGLTSSQAKORLANGLDYNINLDTSPF
YGNITITRTYERLDLTLEKVDAGIKVCSGIVGIGETVDRAGLLQDLNLPPEP
VPINMLVKVGTPLADNDVDFPIKTIQAVARIMPTSVYRISAGREONQTOAMC
PMAGANSIFGCKXLLTTPNPEDEKDLQFLKLGUNPQGTAVLAGDNBQQRRLQALMT
PDTPDEYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/standard_name="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/transl_table="MATVKKOATIAAAGRAAAHYEQHADLQROSDALLMLPORKYT
HYLDAGCGPMMSSRHMRERHAQVTAALDSPMLVQAKQKADADHYLAGDIESPLATA
TFDLAMSNLAVQWCGNLSTALRELYRVVRPGVVAFTTLVQSLPELQAWQAVDERP
HANRFLPPDEIEOSLNGVYOHIIQPTLIMFDDALSMRSIKGIGATHLHGROPRIL
TRSOLORLQAMPQOQGRPLTYHLFGLVARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE-8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/transl_table="MTTDLAFDORHIMWPTSMSTPLPVYVVSABECCELLSDGR
LYDGMSSWMAALHGYNHQNLNAMKSOJDAASHWFGITAPAIELCRKLVAATPOP
LECVLADSGSAVEVAMKMLQYQAKGEARORLFRNGYHGDTFGAMSVCPDMS
HMSLKGYLPEMLKRIKICRREGILLADEIRATGFGTCKLPACNAETADILGCKAL
GMRKYPENLAFAPAROSYMDGEMDRDWWGFARLMAARHETAAVITPEIYOGAG
TGGVITLSTVITREVAETISNGEAGCTPHGTFNGNPLACAAANASLAIIESGDMO
GVADIEVOLREQLAPARDAEMVADVRLGAIQVETTHPVNNAALQKFFVBOGWIRP

```

```

RBS      5088..5100      /standard_name="ORF1 RBS"
gene     5098..5574      /gene="ORF1"
CDS      5098..5574      /gene="ORF1"
          /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
          /codon_start=1
          /transl_table=11
          /evidence=experimental
          /product="PROTEIN"
          /protein_id="CAB69593.1"
          /db_xref="GI:6741866"
          /translation="MKLISNDLRGDKLRPHRVNMGVYDGDNI SPHLAMDVPAGTK
          SFVVCYPPDAPGSGMRHWVYNLPADTRVLPQGSGLVAMPDGVLTQRTDGRKG
          YDGAAPKGETHRITFTVHALDIERIDVDEGASGAMGFVHFSLASITAFMS"
          5583..5605
          5583..5644
          /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
          TERMINATOR"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 ggcgcagcagtgatcgccagcatt 25
          |||
          195 GCGCAGCAGGTGTCATCGCCAGCATT 219

RESULT 8
LOCUS      A93679      5872 bp      DNA
DEFINITION Sequence 6 from Patent EP0798384.
ACCESSION      A93679
VERSION      A93679.1 GI:6741867
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli.
          Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Escherichia.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch,O. and Brass,J.
TITLE      Biotechnological method of producing biotin
JOURNAL      Patent: EP 0798384-A 6 01-OCT-1997;
          LONZA AG (CH)
FEATURES
source      Location/Qualifiers
          1..5872
          /organism="Escherichia coli"
          /strain="DSM498"
          /db_xref="taxon:562"
          /clone="PRO30A15-9"
          1141..1156
          /standard_name="BIOF RBS"
          1154..2308
          /gene="BIOF"
          1154..2308
          /gene="BIOF"
          /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
          /EC_number="2.3.1.47"
          /codon_start=1
          /transl_table=11
          /evidence=experimental
          /product="KAPA SYNTHASE"
          /protein_id="CAB69594.1"
          /db_xref="GI:6741868"
          /translation="MSWDERINAAIDARRAADALRRRYPVAGAGRWLVADRDQYLN
          FSSNDYGLSHHPDII RAMQGAQBFQIGSGSGSHVGVVHQALEELAEELGLYSRA
          LFIISGFANQAVIAMMAKEDRIADRLSHASLLEASTLSPQLRFAHNVTTLAR

```

```

RBS      3030..3045      /standard_name="B10D RBS"
gene     3043..3753      /gene="B10D"
CDS      3043..3753      /gene="B10D"
          /standard_name="DETHIOBIOTIN SYNTHASE"
          /EC_number="6.3.3.3"
          /codon_start=1
          /transl_table=11
          /evidence=experimental
          /product="DTB SYNTHASE"
          /protein_id="CAB69595.1"
          /db_xref="GI:6741869"
          /translation="MSKRYFVTGTDTEVGKTVASCALLQAAKAGYRTAGYKPVASG
          EKPPESLNSDALALQNRNSLQIDYATVPYFAPSPHITSAGEGRPIESLVMSAG
          LRALAEQADWLVVEGAGWTFELSDTFEADWTOEQLPVLLVGVKLGCIHMLTA
          OVTOHAGTLFAGVAVANDVTPPKRRHAEVYVTLTRMIPAPLGEIPWLAENPENMATK
          YINAEVDASTLGFSTR"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 ggcgcagcagtgatcgccagcatt 25
          |||
          195 GCGCAGCAGGTGTCATCGCCAGCATT 219

```

```

RESULT 9
LOCUS      AR101809      5872 bp      DNA
DEFINITION Sequence 1 from patent US 6083712.
ACCESSION      AR101809
VERSION      AR101809.1 GI:12812607
KEYWORDS
SOURCE      Unknown.
ORGANISM      Unknown.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE      Biotechnological method of producing biotin
JOURNAL      Patent: US 6083712-A 1 04-JUL-2000;
          Location/Qualifiers
          source      1..5872
          /organism="unknown"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 ggcgcagcagtgatcgccagcatt 25
          |||
          195 GCGCAGCAGGTGTCATCGCCAGCATT 219

RESULT 10
LOCUS      AR101810      5872 bp      DNA
DEFINITION Sequence 6 from patent US 6083712.
ACCESSION      AR101810
VERSION      AR101810.1 GI:12812608
KEYWORDS

```

SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
FEATURES
Location/Qualifiers
1. 5872
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN /organism="unknown"

Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pied. No. 1.1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcagtgatccgcagcatt 25
Db 195 GCGCAGCAGTGATCCGCAGCATT 219

RESULT 11
LOCUS AE000180
DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.
ACCESSION AE000180 U00096
VERSION AE000180.1 GI:1786988
KEYWORDS
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 11022)
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PMID 9278503
REFERENCE 2 (bases 1 to 11022)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
3 (bases 1 to 11022)
REFERENCE 3 (bases 1 to 11022)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
4 (bases 1 to 11022)
REFERENCE 4 (bases 1 to 11022)
AUTHORS Plunkett, G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: mark@amherst.gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic

Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated: this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source Location/Qualifiers
1. 11022
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
<1. 130
/note="REP (repetitive extragenic palindromic) element; contains 4 REP sequences"
complement(147. 1430)
/gene="ybhC"
/note="b0772"
complement(147. 1430)
/function="orf: Not classified"
/note="fa27: 98 pct identical to fragment YBHC ECOLI SW:P46130 (300 aa) but contains 127 additional C-terminal residues"
/codon_start=1
/transl_table=11
/product="putative pectinesterase"
/protein_id="AAC73859.1"
/db_xref="GI:1786989"
/translation="MNTFSVSRRLALAFGYLTTRCSSPPQRRSDQTAQCTSSRI LSAKEQNPDQHYFASLPGAANNPSITLPAPDVFVGAGTGTHTTQAVD AAIIRKTRKROYIAVWPEYOGTYVPAFGITLYGGEKPIDVKIISLDGNSPA DMRHDVNPGRKYMGPKAPMYGSCSKRSDISGVLSAWEVSQNNGLQDLTENT LGSVDVAGNHPAVALRTDGDVOIINNVTILGRONTFFVNSGVNRLETRNPRLVT NSTIEGDNDIVSGCAVYVDNTEPRVNSRNGOEAIVFAPATLSNIYYGLAVNSRFN AFEDGYAQLGRSLDNDANTNGOVVIRDSAINGEFTAKPMADAVISNRPFRAGTCSVD DNDIEQRNLNDTNYNRMWEYNNRGVSKVAEAKK"
complement(1507. 1535)
/note="factor Sigma70; predicted +1 start at 806574"
complement(1582. 2058)
/gene="ybhB"
/note="b0773"
complement(1582. 2058)
/gene="ybhB"
/function="orf: unknown"
/note="f158: 99 pct identical to YBHB_ECOLI SW: P12994"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73860.1"
/db_xref="GI:1786990"
/translation="MKLISNDLRDGLRPHRVFNGMGYGDGNIISPHLAMDDVPAGTK SFVYVCDPAPRGSGMHWVYVNLPAADRVLVPOGSGCLVAMPDGLQTRDFGKTG YDAAPPKETHTHYIFTVVALDIRVDVDEGASGAMVGTFHFLASASIVAMS"
complement(2108. 2124)
/note="central position to predicted promoter: 85.5"
/bound_moiety="Rhas predicted site"
complement(2117. 3406)
/gene="b10A"
/note="b0774"
complement(2117. 3406)
/gene="b10A"

```

/EC_number="2.6.1.62"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
/codon_start=1
/transl_table=11
/product="7,8-diaminopelargonic acid synthetase"
/protein_id="AAC73861.1"
/db_xref="GI:1786991"
/translation="MTTDLAFDQRIHIMPTSMPLPVVPSAGCEGELLISDGR
LYDGMSSWMAIIGHYNHPTDLNAAKMSOIDAMHVMGEGITIAPALEICRKIVAMPPOP
LSCVFLADSGVAEVAEMKALQYMAKGEARORLITRNCYHGTFCGMSQCDNS
MSIMKGYLPENLAPAPQSRMDCEMDERDMYGRALAAHREITAYITETVAG
GMRKHPKRLKRIKICDREGILITADELTGFGTRGLFACENLEIADILICGKAL
TGGMTLSATLTREVAETISNGEAGCEMHGTFMGNPLACAAANASLAIESGDMQO
QVADIEVQLAPARDAEMVADVGVGALGVETTHPVMAALQKFFVEGVWIRP
FGKILYMPPIYILPQOLRLTAANRAVADTFPCQ"
/complement(2193..2221)
/gene="bioA"
/note="factor Sigma70; predicted +1 start at 807260"
3411..3450
/note="central position to blob promoter: -20"
/bound_moiety="B10b documented site"
3411..3450
/note="central position to predicted promoter: -20; genetic
evidence for the site"
/bound_moiety="B10b predicted site"
complement(3411..3450)
/note="central position to predicted promoter:50"
/bound_moiety="B10b predicted site"
complement(3411..3450)
/note="central position to bioA promoter:50"
/bound_moiety="B10b documented site"
3413..3441
/note="factor Sigma70; promoter bioB; documented +1 at
808525"
complement(3447..3473)
/note="factor Sigma70; promoter bioA; documented +1 at
808515"
3493..4533
/gene="d10b"
/note="b0775"
3493..4533
/gene="d10b"
/EC_number="2.8.1.-"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
/codon_start=1
/transl_table=11
/product="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHRPWTLSQVLELEPKLLDLIFEAQVHROHPDPOVYST
LLSTKAGCPEDCKYCPOSSRYKTLFAERLMEVEYIESARAKAGSTRFCGAAM
KNPHRDMPTLEQWVGKAKGLACMTLGLTSSQAOIRLANGLDYNNILDTSPF
YGNITRTYQERDITLEKVDAGIKVCSGIVGLGTVADKAGLILQLANLPPEPS
VPINMLVKGITPLADNDVDADFEITIVARIIMPTSVRLSAGREXONECTOAC
EMAGANSIFVGCXKLTTPNPEKDKDLFRKGLNPOTAVIACDNEQORLEQALMT
PDEYNAAL"
4530..5684
/gene="bioF"
/note="b10f"
4530..5684
/note="b0776"
4530..5684
/gene="bioF"
/EC_number="2.3.1.47"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAC73863.1"

```

```

Query Match 100.0%; Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ggcgcagcagtgacgcgcagcatt 25
|||||
Db 3571 GCGCAGCAGTGCGATCGCAGCATT 3595

```

```

RESULT 12
AE005258 13501 bp DNA BCT 21-MAR-2001
LOCUS
DEFINITION
Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
of 155.
ACCESSION
AE005258 AE005174
VERSION
AE005258.1 GI:12513751
KEYWORDS
ORGANISM

```

```

Escherichia coli O157:H7 EDL933.
Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

```

```

REFERENCE
1 (bases 1 to 13501)
Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Weich,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)

```

```

TITLE
JOURNAL
MEDLINE
21074935
PUBMED
11206551
REFERENCE
2 (bases 1 to 13501)
Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Weich,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

```

```

FEATURES
source
1..13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
<1..7576
/note="O-island #36; Region of the EDL933 chromosome not
homologous to E. coli K-12 M61655; Cryptic prophage
CP-933K; includes one copy of the 13 bp direct repeat that
flanks the prophage"
66..665
/gene="lomK"
/note="20981"
66..665
/gene="lomK"

```

```

CDS
66..665
/function="putative membrane; Other or unknown (Phage or
Prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to
residues 1 to 199 of 199 from GenPept 118 :
g117532789|gb|AA63231.1|AF151091_2 (AF151091) lom
[prophage P-ElbA]"
/codon_start=1

```

```

/transl_table=11
/product="putative outer membrane protein of prophage
CP-933K"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MRKYCAATLSAATCAVSGVPAMASEHOSTLSAGYLHASTDAPG
SDLLGKIVKRYEFTDGLITFSVANAEDCKTHSDTRHEDDYRANMFPSVAG
PSVRVNEFSAAMAGVAYSRYSSTSGYFRPTDNRKTHVLGSLDARYSNLSLAW
GAGVQNPTEPSAIVADYAYEXSGSGDWRTDGFVGYKAF"
/gene="20981"
/length=719
/gene="20982"
/length=719
/gene="20983"
/length=719
/function="putative structure; structural component (Phage
or Prophage Related)"
/notes="Residues 164 to 440 of 440 are 68.79 pct identical
to residues 381 to 645 of 645 from Genpept 118 :
g1145854361gb|AAD25464.1|AF125520.59 (AF125520) putative
tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="putative tail component of prophage CP-933K"
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAVQISVILKQAGKPIQNCITQKARRNSTXYVNVTVASE
NDEAGRTSMDEYQCYSTLVLEGFPPSHACTIYVEDSOPGTLNDFIGANTEDAR
PALRRFEQWVEEARHAEARKNAGEFSAKRNAGISAKSEASANAADVISAEDABE
SARQAEASASAKSEASASSASAEASASASASASASASASASASASASASASAS
TSTETAREASASASASASASASASASASASASASASASASASASASASASAS
RDTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
TGLTFLDPKDATQVQGMFQHLQYRFGDGMQVQKGLDEVSGTGRTE"
/length=2004
/gene="20984"
/length=2004
/gene="20985"
/length=2004
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 of 129 from Genpept 118 :
g1145854371gb|AAD25465.1|AF125520.60 (AF125520)
hypothetical protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MKWAVIOAELENDMMNLRKLMQSLGCGKHDCEGSRITQQLR
LGPADLLESDENGIIPEDORVITQVILLADAKKQICVVRPIQLIRADGTWENIGMK
"
/length=2489
/gene="20985"
/length=2489
/gene="20986"
/length=2489
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from Genpept 118 :
g1169603671gb|AAF33527.1| (AF170176) hypothetical protein
predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPRTTFPHNSVNIQSSPCQTVSPAKKEYELKVIDEYKPTLLF
QWEPNPERYKADVEIVNTKQHPYLDNTVNAARIESDRMIGIFVDGDSVNGKTAFS
KLEDPENRYMIVREDYVDFSMYKRLSDIYHDIQELRTEREDKRDYELINLEKEIR
EISKAOSLSIMYAKKRNHAWPDPFRNLALTKGELFRCTYNTKNGHSIGCECTYLD
MDKLITKLTITIAPDGISMHVDRNDSDVINESSALITVNSNHPALLEGLSTPHMSKVD
AHPYVGLGKGVKYEFTPLHNTHPCDIETFNHPILIMNTISQYTCSW"
/length=3530
/gene="20986"
/length=3530
/gene="20987"
/length=3530
/gene="20988"
/length=3530
/gene="20989"
/length=3530
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="No significant matches"
/related="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPSLQSNFNSAPAGCYSPAPIAPNNAENAYADYVIDICKRPL
SAADLSNYESVIRAVHDSRSRLIDQHTVMIGNATYVDALSRQTEPDAVSTIGHK
VHICIKRYNEVELNEESSVKIDIDQSLTCENELYEDVGOEPIPTICEAENDEEY
VSFVAPDTPDSWEMPQWOGELIHEIHHYTGSSDPSGDSNIELEGEIILARYAOELG
MSVPDFKGYVAPEPEREHLRLRNALROAAMHRENERAFERLGTISDRYEASDPFT
EYSVSVINGICFIOQHPFGLAINDNLDQANQIOLYHGAFYITFGDYKHNQ"
/length=5350
/gene="20989"
/length=5350
/gene="20990"
/length=5350
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="No significant matches"
/related="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPYSVNIQCSMNSLRLNLTSPDNRYLSSVPAVAVSHNGAOY
KVGNTAVVATVTDKFCVTRSHSGCFTNLRLHKGWKEISRTIEVNLASPSAM
ERGIYHSNRPLPVDYAPLPELVYDNRSLSPVNVIGKGNVAYVADADARKYLKM
FTTOSNEVTESEVCEVNOYAGASAEIYVNGDIDIGRMDKINGSLINISLPQO
AEHALYMDPRLBOGKILFVDTETNVLYDAKKNFPIIDISSVNSVDRSMSEQIMQ
SYHGKQDLISVLSKI"
/length=6462
/gene="20990"
/length=6462
/gene="20991"
/length=6462
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="No significant matches"
/related="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPISLNVILHOSRSSMSDPTIESLYKASVQWIKRNPQLRF
OCTDNITVOOIEALDKIGSTETGVLINAIESTSRKSEYVYHLLNSRLGVANHD
IDAENHRTGSDPHCNLNAVYPCGEGISVYDFATTIVFHLHVFHNLNGERLVES
SRAESQKYSPLLEAFRTVGLGARFSEVLSENKFEELGMBRRTSYPSXALIHDDNT
VSLGFOOYRLHPL"
/length=6462
/gene="20991"
/length=6462
/gene="20992"
/length=6462
/notes="20992"
/complement(7668)
/gene="ybh8"
/complement(7668)
/function="orf; Unknown function"
/notes="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDQDKLPHRIVFNGKGYDDDNISPHLANDVPAGTK
SFVVCYDPAITGSGMMWVWVNLPLADTRVLPQGFSGGLAAMPDGVLOTITDCKTG
YCGAPPKKETHRVYFTYHALIDVDEGASGAMVGFNFHSLASASTAMPS"
/complement(8203)
/gene="b10a"
/notes="b10a"
/length=3530
/gene="20993"
/length=3530
/complement(8203)

```

Query Match 100.0%; Score 25; DB 1; Length 13501;
 Best Local Similarity 100.0%; Pred. No. 0.977;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcatgcagcagcatt 25
 |||||||
 Db 9657 ggcagcaggtgcatgcagcagcatt 9681

RESULT 13
 LOCUS AP002553 297816 bp DNA BCT 07-MAR-2001
 DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
 ACCESSION AP002553
 VERSION AP002553.1 GI:13360211
 KEYWORDS
 SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub-strain:RIMD 0509952)
 DNA.

ORGANISM Escherichia coli O157:H7
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE
 AUTHORS Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S.,
 Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
 Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
 Sasaki, C., and Shinagawa, H.
 Complete nucleotide sequence of the prophage VT2-Sakai carrying the
 verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
 derived from the Sakai outbreak
 Genes Genet. Syst. 74 (5), 227-239 (1999)

JOURNAL MEDLINE
 REFERENCE
 AUTHORS 2 (sites)
 Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
 Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
 Hayashi, T.
 Comparative analysis of the whole set of rRNA operons between an
 enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
 Escherichia coli K-12 strain MG1655
 Syst. Appl. Microbiol. 23 (3), 315-324 (2000)

JOURNAL MEDLINE
 REFERENCE
 AUTHORS 3 (sites)
 Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
 Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
 Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C., and
 Shinagawa, H.
 Complete nucleotide sequence of the prophage VT1-Sakai carrying the
 Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
 O157:H7 strain derived from the Sakai outbreak
 Gene 258 (1-2), 127-139 (2000)

JOURNAL MEDLINE
 REFERENCE
 AUTHORS 4 (sites)
 Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
 Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
 Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasaki, C.,
 Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
 Shinagawa, H.
 Complete genome sequence of enterohemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12
 DNA Res. 8 (1), 11-22 (2001)

JOURNAL MEDLINE
 REFERENCE
 AUTHORS 5 (bases 1 to 297816)
 Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
 Hayashi, T.
 Direct Submission
 Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
 Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
 Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
 URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
 Fax: 81-6-6879-2047)
 genome project.

COMMENT
 FEATURES
 SOURCE Location/Qualifiers
 . 1..297816
 /organism="Escherichia coli O157:H7"

gene
 CDS

gene
 CDS

gene
 CDS

gene
 CDS

/strain="O157:H7"
 /sub_strain="RIMD 0509952"
 /db_xref="taxon:83334"
 79..1245
 /gene="ECS0753"
 79..1245
 /gene="ECS0754"
 79..1245
 /gene="ECS0753"
 /note="similar to SUCC_ECOLI g111786948 percent identity
 100 in 388 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="succinyl-CoA synthetase beta subunit"
 /protein_id="BAB34176.1"
 /db_xref="GI:13360212"
 /translation="MNLHEYOAKOLFARYGLPAPVACTPTPREAEASKIGAGPW
 VKCOVHAGRGKAGGVKVVNSKEDIPAFAPENMLGKRLVYOTDANGOPNOLVSEAT
 DIKELYLCAVNDRSSRYFMASTGEVIEKVAEETPHLIRKALDPVTPMRYQG
 RELAFKLGEGKLVQOFTKIFMKLATIFLERDLALIEIPVITVQGDICDGLGA
 DGNALFRQDPLREMRDQSEDPREAOAQMELNYVALDGNIGCMVNGAGLAMDITV
 KLHGEPANFLDVGAGTKERVEAFKIIISDVKYAVLVNIFGIVRCDLIADGIIIG
 AVAEVGVNVPVVRLEGNNAEIGAKKLADSGINIITAAKGLTDAQOYVAAYVGK"
 1245..2114
 /gene="ECS0754"
 1245..2114
 /gene="ECS0754"
 /note="similar to SUCC_ECOLI g111786949 percent identity
 100 in 289 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="succinyl-CoA synthetase alpha subunit"
 /protein_id="BAB34177.1"
 /db_xref="GI:13360213"
 /translation="MSILDKNTKVIYCOFTSGQTFHSEDAIAYGTRMGVTPFGK
 GTTHLGLPVNTVREAVATGATASVIVYPAFCDSILIEADIKILITTEBTP
 LMLTLKVLDEAGVMIIGPNCGVITPGECKIGIOPHIKPGKVGIVSRGLTVE
 AVKQETDVGEGSTCGDIPGSDPILEMEKRPQREIYALMIEIGSAAEEA
 AAYIKENYKRPVYCIAGVTAPKGRMKAGAIITAGKGTADKFAALEAGVTVRS
 LMDIGALKTVLTK"
 2781..3686
 /gene="ECS0755"
 2781..3686
 /gene="ECS0755"
 /note="probable transcriptional regulator, similar to
 transcriptional activators e.g. glycine cleavage system
 transcription activator (gcv operon activator) -
 Escherichia coli g14170431sp|P32064|GCV_A_ECOLI percent
 identity 31 in 300 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="putative transcription regulator"
 /protein_id="BAB34178.1"
 /db_xref="GI:13360214"
 /translation="MRGKIPKTELLVPEVVARHESYTRAEEALNQSAAVFRVSAL
 EEFLTPLEFSKKRIFLNDGKYVIGIVETLKKLRDPTITMTQNPQVDELAN
 PFTSHMLIPNHETKRLPDITVINSLANGPFLAREYAVITRENFCAPAEVEY
 LFEEBILPVCSSILAMSDOKLSVAELTELPLDHSRTITGWEWALSGVSSPLVN
 NRPDLDSMLIAAVRSNLGVALLPRAIIOHDDSGMVIPCDVPIRTGNRFMTWOB
 EKSDSPHIOQFREWMLAKSVVQEM"
 complement(3720..4332)
 /gene="ECS0756"
 complement(3720..4332)
 /gene="ECS0756"
 complement(3720..4332)
 /note="probable cob(1)atamin adenosyltransferase, similar
 to cob(1)atamin adenosyltransferases (corrinoid
 adenosyltransferases) e.g. [Escherichia coli]
 g11151481sp|P13040|BTUR_ECOLI percent identity 67 in 200
 aa"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental

```

/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTEHHVROOKLEQVDTRVAAAEIKGILLIVTNGNGK
KSTAFGTVTAAGKTVGAQYIKQMDNGEVLQPLGEFHIMGTFTWTONR
QADIDAAKEVSESKRMADKRYDLVLDLTYMLAYHLYLTERVIASLQRPQOOSV
ITVGGCHSOLIKMADYSELRPVKHAFTDNGIQAPGIDW"
/complement(4332..5984)
/gene="ECS0757"
/complement(4332..5984)
/gene="ECS0757"
/translation="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
g1120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/translation=1
/evidence-not experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPFIMQEPFLQNKDTEYTLISDQIITVELDGEVIAKPAE
ATLLARQAFYASFFLSAHLQVYASLNDQASNDKYALQLIRNAEVSAGVLP
NCODGTATIVASGQOLWGTGNDALSKGTSTFQFNRLRFSQNALPDMTEVNTQ
TNLPNOIDISAGDEYHFLCNKGGCSANKAALYQETKSLQPEKILTAPEKMSL
GTACPPYHIAFVGVGSLSDDTLKTAKLSTRYDNLPTSCGSGQAFRDLTEKVL
EASQGFQGAQFGKRYFADILRVILPRHSGSCTPAMALSCSADNIAKINKGIML
EKLEHNPQOYIPASLRBNHQAHOVLDIRLPVGMODLARLPVGTSLGPIVVAR
DIAHAKIKARLDGEMPEYKIHIVYYPAPKTPENMACSISGPTTGRMDGYIDTF
QAGSGLVLSKGRSQQVTDACHHGGENGIGGAALLAOEYVKSLRCLVEPELG
MEAVMMVEYENPAFLVDKGNFESQPEQHRCASCPACH"
/complement(6092..7372)
/gene="ECS0758"
/complement(6092..7372)
/gene="ECS0758"
/translation="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g1121467|sp|P24943|GLUT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/translation=1
/evidence-not experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLTMTLALAVLGMIGVLTNTASPETAKLYAOEISIFT
TIFLRILKIMILAVLSTLVGIAKGDARLGRIFESKTLFLFCASLLSLALGLITV
NFFMGGTGTFNFAHGAETTGVAAREPFTKXFIASHAPSTIVDAMAHNBILQIVESI
PLGCSITAIKGGSAIVHALDSLAAHMLKLTGYMLPAPLYFPAALISALIRGALAV
VSAGITPGEFTFTMLLWVLLGLAIVYGPICIRKTRIALSPALLFTTSSEAPF
GTLEKLEQFVSPIKIASFVLPISFNLSMAYSFAVFIQAQCNHLSIGEQITM
LILMLTSGKAGVPRASMVVIAATLNDPFIPEAGIILMGVDPFLDMGRSATVMSN
AAGAAMVSRMEGHEGCGRCALKKPESNALP"
/complement(7533..7850)
/gene="ECS0759"
/complement(7533..7850)
/gene="ECS0759"
/translation="unknown"
/translation=1
/codon_start=1
/translation=1
/evidence-not experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

```

Query Match 100.0%; Score 25; DB 2; Length 297816;

Best Local Similarity 100.0%; Pred. No. 0.63;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 92389 GCCGACGAGTGATCGCCAGCAT 92413

```

RESULT 14
E00893
LOCUS E00893 1121 bp DNA PAT 29-SEP-1997
DEFINITION Genomic DNA encoding biotin synthetase.
ACCESSION E00893
VERSION E00893.1 GI:2169154
KEYWORDS JP 1986149091-A/1.
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1121)
AUTHORS Hirono,Y., Kojima,T. and Kimura,H.
TITLE DUPLICATION OF BIOTIN
JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
COMMENT NIPPON SODA CO LTD
OS Escherichia coli
PN JP 1986149091-A/1
PD 07-JUL-1986
PF 24-DEC-1984 JP 1984272605
PI HIRONO YOSHIIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N1/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
C12R1:19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=Escherichia coli Ns101;
CC feature is identified by experimental;
FH Key Location/Qualifiers
FT CDS 42..1079
FEATURES
source /product="biotin synthetase".
location/Qualifiers
1..1121
/organism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT 289 a 296 c 325 g 211 t
ORIGIN
Query Match 93.6%; Score 23.4; DB 10; Length 1121;
Best Local Similarity 96.0%; Pred. No. 6.9;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
DB 120 GCCGACGAGTGATCGCCAGCAT 144
RESULT 15
AF250776
LOCUS AF250776 5526 bp DNA BCT 31-JAN-2001
DEFINITION Uncultured bacterium pCosHE2.
modc-bioA intergenic region, DAPA-antitransferase bioA (bioA),
biotin synthetase bioB (bioB), KAPA synthetase bioF (bioF), and
biotin biosynthesis protein bioc (bioc) genes, complete cds, and
dehbiobiotin synthetase bioD (bioD) gene, partial cds.
ACCESSION AF250776
VERSION AF250776.1 GI:12620124
KEYWORDS unclutred bacterium pCosHE2.
SOURCE uncultured bacterium pCosHE2.
ORGANISM Bacteria; environmental samples.
REFERENCE 1 (bases 1 to 5526)
AUTHORS Entcheva,P., Liehl,W., Johann,A., Hartsch,T. and Streitz,W.R.
TITLE Direct cloning from enrichment cultures, a reliable strategy for
isolation of complete operons and genes from microbial consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE 20575196

```

PUBMED 11133432
 REFERENCE 2 (bases 1 to 5526)
 AUTHORS Entcheva, P., Liebl, W. and Streitt, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
 FEATURES
 source Location/Qualifiers
 1..5526
 /organism="uncultured bacterium pCosHE2"
 /db_xref="taxon:143797"
 /clone="pCosHE2"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(52..528)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical 17.1 kDa protein in modc-bloc
 intergenic region"
 /protein_id="AAG60577.1"
 /db_xref="GI:12620125"
 /translation="MKLISNDLRGDKLPHRHVNGMGYDGNISPHLAMDVPAGTK
 SFVYCYDPAPTGSGMWHVYVNI.PADTRVLPGFGSLVAMPDGYLQRTDFKAG
 YDGAAPKGETHRYITVHALDIERIDVDEGASGAMGVNHFHSLASSTAFMS"
 complement(587..1876)
 /gene="bioc"
 complement(587..1876)
 /gene="bioc"
 /note="7,8-diaminopelargonic acid
 synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase BiocA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDDLAFDORHIMPYTSPPLVYVSAEGCELLISDGR
 LVDKSSWMAIHGYNHPOILAAKMSQIDAMSHVFGGITHAPAEICRKIVAMTPOP
 LKCVLADSGVAEAVAMKMLQYQAKGEAROLFERNHGDTEDGASVCPDPS
 MSLMGKYLLENLFAAPASRMGEMDERDVGARLMAHRRHIAVITTEPVQAG
 GMRVHPEMKRIRKICDRGILIIADPIATGERTRKLCACAEIAPILICGKAL
 TGTGNTLSATITTEVATTSNCFAGCFMGSPYTWGNPLACANASLALIESGDWQ
 OVADIEVQLREQLPARDAEVADVRLGAGVETTHPVNMAALOKFVEQGVWRP
 FGKLIYMPYIILPQOLRLTAANRAVODEFFEQ"
 1963..3003
 /gene="biob"
 1963..3003
 /gene="biob"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase Biob"
 /protein_id="AAG60579.1"
 /db_xref="GI:12620127"
 /translation="MAHRPMWTLISQVLELPEKPLDLLFEAOVHROHPPOVOST
 LLSITKGACPEDCKYCPOSSRYKTGLEAELEMEVOYLESARAKAAGSTRPMGAW
 KNPERDMPLTEWOGVCKDGLACHTLTGLTSSQRLANGLDLYNNIDTPEF
 YGNTITRTIOERLDLEKYDAGIKVCSGIVGLGSCQRLANGLLDLYNNIDTPEF
 VPINMLVKGKGLPLADNDVADDFITTIIVARIIMPTSVYRISAGSENEQTOAC
 FMAGANSIFYGCKLITTPNEBEDKDLFRKGLNPQOATVLAGDNEQOQLBOALMT
 PDDEYYNAAL"
 3000..4154
 /gene="biop"
 3000..4154
 /gene="biop"
 /note="8-amino-7-oxononanoate synthase"
 /codon_start=1
 /transl_table=11
 /product="KAPA synthetase Biop"
 /protein_id="AAG60580.1"
 /db_xref="GI:12620128"
 /translation="MSWQEKINAALDARRAALRRRYPAVQAGHMLVADDROYLNF
 SSNDYGLSHHPQIIRAMQOAGDFGSGSGSVASGYVHQALEDEELAEMLGYRA
 LFIISGAANOAVIAAMAKEDRIADRLSHASILFAASISPSQLRRPFIANDYTHAR
 LLASPCGOQLVTEGVFSMDGDSAPLAEIQOVTQOHNGKMLMDAHDGTVIGEGSGR

gene
 CDS
 4141..4896
 /gene="bioc"
 4141..4896
 /gene="bioc"
 /note="reaction step prior to synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=11
 /product="biotin biosynthesis protein Bioc"
 /protein_id="AAG60581.1"
 /db_xref="GI:12620129"
 /translation="MATVKNQALIAAFGRGAALIEQHADLQROADVLLAMPKRYT
 HYLDACGCGMTTMRERHRAQYALDLSPPMLVOAROKADADHYLAGDIESTPLATA
 TFDLASNLAVQMGNI.STALRELTVSVSKGVAVETTLVQSGI.PELIQAOVADERP
 HANRFLPDEIEQSLNGVYOHHLIOPITIMPDALSAMSLKIGATILHESDRRL
 TRSOLQRLAMPQOQGRPLTYHFLVIARE"
 4883..5526
 /gene="biob"
 4883..5526
 /gene="biob"
 /note="DTP synthetase"
 /codon_start=1
 /transl_table=11
 /product="dethiobiotin synthetase Biob"
 /protein_id="AAG60582.1"
 /db_xref="GI:12620130"
 /translation="MLVSKRYFTVGTDFEVCRTVASCALLQAKAAGRYTAGKPYAS
 GSEKPEGLRNSDALALORNSSLQIDYAVNVNTPAEPSPHILISAOGRPRESSVMS
 SGIRALEQOADMVYEGAGCFWPSLSDPTTFEDWYQOLPYIVVYKGLGGINRAML
 TAOAIDHSLTLAGVYANDVTPGKRHAETITTLTRMIPALIGELPWLAENPE"
 BASE COUNT 1274 a 1507 c 1567 g 1178 t
 ORIGIN
 Query Match 93.6% Score 23.4; DB 2; Length 5526;
 Best Local Similarity 96.0% Pred. No. 5.5;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ggcgacgagtgatcgacgacatt 25
 Db 2041 GCGCAGCAGGTCATCGCAGCAT 2065
 RESULT 16
 AE004895
 LOCUS 12211 bp DNA BCT 30-AUG-2000
 DEFINITION Pseudomonas aeruginosa PA01, section 456 of 529 of the complete
 genome.
 ACCESSION AE004895 AE004091
 VERSION AE004895.1 GI:9951083
 KEYWORDS
 SOURCE
 ORGANISM
 Pseudomonas aeruginosa.
 Pseudomonas aeruginosa
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 1 (bases 1 to 12211)
 REFERENCE
 AUTHORS Stoyer, C.K., Pham, X.O., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
 Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M.,
 Garber, R.L., Goltz, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y.,
 Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,
 Smith, K., Spencer, D., Wong, G.K., Wu, Z., and Paulsen, I.T.
 TITLE Complete genome sequence of Pseudomonas aeruginosa PA01, an
 opportunistic pathogen
 JOURNAL Nature 406 (6799), 959-964 (2000)
 MEDLINE 20437337
 REFERENCE
 AUTHORS 2 (bases 1 to 12211)
 Stoyer, C.K., Pham, X.-O.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
 Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,
 Lagrou, M., Garber, R.L., Goltz, L., Tolentino, E.,
 Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
 Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,

TITLE
JOURNAL

Wong, G. K. S., Wu, Z., Paulsen, I. T., Reizer, J., Saier, M. H., Hancock, R. E. W., Lory, S. and Olson, M. V.
 Direct Submission
 Submitted (16-MAY-2000) Department of Medicine and Genetics,
 University of Washington Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA

FEATURES

Source

Location/Qualifiers
 1. .12211
 /organism="Pseudomonas aeruginosa"
 /strain="PA01"
 /db_xref="taxon:287"

gene

CDS

109. .1344
 /gene="PA4816"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG08201.1"
 /db_xref="GI:9951084"

gene

CDS

1320. .1814
 /gene="PA4817"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG08202.1"
 /db_xref="GI:9951085"
 /translation="MAGPSALFLTARTSATCAQPLIRARPPRTAPPKIPLAAD
 PRLDITVVKYRGLCECNATCTTLPVREIDLLIMRLVDEPEREPKRIAKREK
 DSVTEHFNHRELIFTLRMANSGCLYLDKRLCTTYARRPDTGRNHPRIAGPPTCA
 YRPR"

gene

CDS

complement(1817. .3253)
 /gene="PA4818"
 complement(1817. .3253)
 /gene="PA4818"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAG08203.1"
 /db_xref="GI:9951086"

gene

CDS

complement(3250. .4224)
 /gene="PA4819"
 complement(3250. .4224)
 /gene="PA4819"
 /codon_start=1
 /transl_table=11
 /product="probable glycosyl transferase"
 /protein_id="AAG08204.1"
 /db_xref="GI:9951087"
 /translation="MTPEPLASKLEAPCLLSYVVPVNEAVIPAFHRLGAVIGDL
 GDDEVLYVDGSDRSNAIILAOQADRRYGAFTFRNGKEGAMSGALSGAAY
 IVIDADLDPELLPOMLDAMIIGAEEVNMKRSRAESWLKRSASFYKLRILSE
 VPIPRVDGDFRLRRVVDALCELPENRPMKGI.FAMVGFROVDIAYSRAVAGHSK
 WRVRLNMFALGIGTGSTAPLKVAGYLGILLALATLILAGIFGQOEIHEHPVIA

gene

CDS

ALFLIGGLIOLAIQVCSSEYLGRMTMEARRRPLVLDIYLPQVGGANP"
 complement(4221. .4583)
 /gene="PA4820"
 complement(4221. .4583)
 /gene="PA4820"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG08205.1"
 /db_xref="GI:9951088"

gene

CDS

4820. .6181
 /gene="PA4821"
 /codon_start=1
 /transl_table=11
 /product="probable transporter"
 /protein_id="AAG08206.1"
 /db_xref="GI:9951089"

gene

CDS

complement(6513. .8180)
 /gene="PA4822"
 complement(6513. .8180)
 /gene="PA4822"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG08207.1"
 /db_xref="GI:9951090"

gene

CDS

complement(8223. .8432)
 /gene="PA4823"
 complement(8223. .8432)
 /gene="PA4823"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG08208.1"
 /db_xref="GI:9951091"

gene

CDS

complement(8483. .9259)
 /gene="PA4824"
 complement(8483. .9259)
 /gene="PA4824"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG08209.1"
 /db_xref="GI:9951092"
 /translation="MTLHRRPAPPLFOPPLPALRRALSRLGFEVAAALCYLLAL
 SEKRAAGATRVVDGSDIYVGEENLDAMQDREHNRHNGQVLSLSPCFPSALPSYDLS
 AALVRGDAGRHSRLSPELKTPPLARDDLGLALAFALSADLYLDRRAAFEGAGRNLP

ISYQPEALRLNAGVGLGHAYAEGERRRHMMWGMEXYRVQ3PILTLIAERFGESAGDS
GMOAGPRHLRLGRLDLVLGHLRGERERMLVSGATLRF"
/gene="mgta"
/complement(9357. .12068)
/note="PA4825"
/complement(9357. .12068)
/gene="mgta"
/codon_start=1
/transl_table=1
/product="Mg(2+) transport ATPase, P-type 2"
/protein_id="AAG08210.1"
/db_xref="GI:951093"

/translation="MSVATQRTQIDTERNDRLSMRAREAONSLETTLTNVRGROD
GLDQEAARLHADGPNVADHPHVLVOLLQAFNNPFIYVLTLSAISFTFEMLP
LDGSETDLTGIVILCVLVSLGMLRMOEXRSGAAALKAAMNTATVLRBERG
RGELREIPMGLVGDIVLISAGDMIPADIRLIESRDLFSOAVLTGALPVEKYDIL
GAVREKSAKRIADQDDLELPTICFMGTNVSGTATVAVATGARTYFESLASIYG
SRAOTAFRGNVSNWILLIRFIMVWPVILLINGFTKQMTAEAFALAVAGLTFEP
LPMYTSANLAKAVAMSRKVVYKRLAIQNGFADVICDTKGTITQDRILIEHVAD
VAGRCDDVLOAMUNSFHSGMKMLMDRAVAEONPRITPSDAMKADDELPEFV
RRRLSVILLEGDQEQILVCKGAVEEMELATRVHODGVDLPLDEERRALLALEQV
RDGFVLLIGTRLSRTESQAYGASDERDVLIAGLTFELDPKETAGPATAALRENG
VAVKVLTDNPNVSAKICREYGLDVGEPDLGRDIDLMODATLORLAEERVFATLPL
OKSRVLAQNGHGVFGINDAPALRADVGISVDSGTDIAKESADILILEKSL
MYIEGVKIGRETCNTMKYLMTASSNFGVSVYASATIPPLMATHLLQNLN
YDISLSPMDKMEPLAKRKDKSKIGFHWIGTSSITPITTYALMWEFAN
SPMOSLTFQSGMTEGLSTQTLVHMLTOKIPFIQSTALPVMALMTGLVWALGIVYP
FSPGAVVGVPLPEFPMVLVGLLCCVNAQTMKTLIRRFQGM"

Query Match 81.6%; Score 20.4; DB 1; Length 12211;
Best Local Similarity 95.5%; Pred. No. 1e+02; Mismatches 1; Indels 0; Gaps 0;
Matches 21; Conservative 0;

OY 3 gcagcagtgatcgccagcat 24
|||||
Db 9863 GCAGCAGTGATCGCCAGCAT 9884

RESULT 17
LOCUS RICOSA1 3053 bp mRNA PLN 12-SEP-1998
DEFINITION Oryza sativa OSA1 mRNA for H-ATPase, complete cds.
ACCESSION D10207
VERSION D10207.1 GI:218178
KEYWORDS H-ATPase; OSA1: hydrolytic enzyme.
SOURCE Oryza sativa (strain: Nipponbare K-1) callus cDNA to mRNA.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 3053)
Wada, M., Takano, M. and Kasano, K.
Nucleotide sequence of a complementary DNA encoding plasma membrane
H+-ATPase from rice (Oryza sativa)
Unpublished (1992)
Data kindly submitted in computer readable form by: Wada Masato
Institute of Biological Sciences
Tsukuba University
Tsukuba City
Ibaraki, 305
Japan
Phone: 0298-53-4871
Fax: 0298-53-6614.
FEATURES
source Location/Qualifiers
1..3053
/organism="Oryza sativa"
/strain="Nipponbare K-1"
/db_xref="taxon:4530"
/cell_type="callus"
31..2901
/gene="OSA1"
31..2901
/gene="OSA1"

/EC_number="3.6.1.3"
/codon_start=1
/product="H-ATPase"
/protein_id="BAA01058.1"
/db_xref="GI:218179"
/translation="MAEDKGLDAVLKESVDLENIPIEEFONLKCRCROGLSEEAOL
RLQFGPNKLEEKESKFLKFLGFMNPNLSWWEALMALAKRAKLRNGSTEEBALVGDITSI
ITILLINSTRIFEEENMAGNAALMALAKRAKLRNGSTEEBALVGDITSI
KIDITIPADARLLEGPPLKIDQSLTGESLPATKGGDVSSTVKGELAVAT
GVTFEFGKAAHLVDSTNQGVRQVLTALGNCICSI AVGMVEIIVMPYIQRVYR
GIDNLLVLLIGIPAMPVLSVTMAIGSHRISQGAITRKMTAIEEMAGMDVLSK
TGTLTLNKLTVDNKLIIEFERGTQDOVILMAASRTENODAIDTALGMADPKEA
RAGIOEVHLEFPNPTDKRTALFYIDSGMKYRSGAPDOIINLHNTQIRRVHAV
IDKEAFGLRSRLAAVQVDEGRKSGPGPMPFVALPLPDPDRHDSATIRRALNG
VNMKRTIGDOLAIKRETRIGRIGCMGTMYSSALQCNKRESVALPYDLIRKAGFA
GVPEPKETVYKRLQARKHICGHTGQGVNDPPLKADIGIIVADATDARSASDVL
TEPGLSVIISAVLTSAITRAIFQRMKNYTVAVSITIRIVFEMLLAIWEDPPEFVLI
IALINDGTIMTISKDLVKPSPLPDSWKLAIEFTTGVLGGLYAMMTVSSGLHPTPT
SLGFTSKALRLQLODDYOKLASAVYIQVSTISQALIFVTRSRSMFIERPGLVFA
FEVAOLIALTIAVYAMMAFTSTIKIGMGAGIWMYXNLVYFPLDIKFLIRYALSGK
AMDVLEORLAFTRKKDPKREPERKMAHARTLIGLOPPDKKPPPERTGYSELNOMA
EEKRAEALRLRLHTHTLKGHVESYKRLKGDIDITIHOSITY"

Query Match 80.8%; Score 20.2; DB 14; Length 3053;
Best Local Similarity 88.0%; Pred. No. 1.5e+02; Mismatches 3; Indels 0; Gaps 0;
Matches 22; Conservative 0;

OY 1 gcgcagcagtgatcgccagcat 25
|||||
Db 412 GCAGCAGCAGCATGCGCAGCAT 388

RESULT 18
LOCUS AC079633 123926 bp DNA HTG 10-FEB-2001
DEFINITION Oryza sativa clone OSJNBa0032608, *** SEQUENCING IN PROGRESS ***
ACCESSION AC079633
VERSION AC079633.8 GI:12739739
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
Oryza.
1 (bases 1 to 123926)
McCombie, W.R.
Rice genomic sequence
Unpublished
2 (bases 1 to 123926)
McCombie, W.R.
Direct Submission
Submitted (06-SEP-2000) Lita Annenberg Hazen Genome Center, Cold
Spring Harbor Laboratories, 1, Bungtown Road, Cold Spring Harbor,
NY 11724, USA
On Feb 10, 2001 this sequence version replaced gi:11225373.
COMMENT
----- Genome Center
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
----- Project Information
Center project name: OSJNBa0032608
Center clone name: OSJNBa0032608

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

TITLE
 Journal
 Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1. 12891

FEATURES

source
 /organism="Vibrio cholerae"
 /strain="N16961"
 /serotype="O1"
 /db_xref="taxon:666"
 /note="biotype: E1 Tor"
 complement(84. .1370)
 /gene="VC1111"
 complement(84. .1370)
 /gene="VC1111"
 /note="similar to GB:J04423 SP:P12995 PID:457106
 PID:490217 GB:000096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="adenosylmethionine-8-amino-7-oxonanoate
 aminotransferase"
 /protein_id="AAF94270.1"
 /db_xref="GI:9655582"
 /translation="MMDLAFDRCHIMHPTSTLPTICYPVSADGVMTLEDGRL
 VDGMSSMAATIHGYNHFKLNQANQJTEQSHIFGGTLHPALICQKLTLPANM
 QOVLADSGSVAVESSKMLQYHNSGQPRAPKFLTRHGHGDTFAMSVTDGNSM
 HSLKGFLEPHIFANSPEGCFDAMDERDIADFRHKLLENHNOIAVALLEIYDAGG
 MRYHPEFLRQVRAICDEFVGLILDELATGFGTGKLFACENAGIOPDLICVAKAL
 GGYMTLSATLTQQOAVDVCAGACGFCMHGPTFMGNPLACVASASLILEGCGKQ
 VANLEAFPEQLPRLNDSRYKOTRMIGALIGVETHLPVNNEVIGALFVERGWIRF
 GLIIMPRYSIOBEHIQKLVDAAMAAALNPDCRQ"
 1497. .2549
 /gene="VC1112"
 1497. .2549
 /gene="VC1112"
 /note="similar to GB:J04423 SP:P12995 PID:145425
 PID:490219 GB:000096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase"
 /protein_id="AAF94271.1"
 /db_xref="GI:9655583"
 /translation="MEVRHNMVTAEVKALDKPMDLFEAOOVHRLHPHNVOST
 LSTKTCACPEDCYCPQSAHYRTVDKERMLEVERYLDAQAKKNGSTPFCGAM
 KNKERMPLKEMIRGVKDMGLCTMTLGLTTPDOQLAQAGLDYNNIDTPSEF
 YGNITRTTIDRLDTLSHVADGKMLCSGIIIGMGESTNDRAGLLLELAINPHES
 VPINMLVKVGTPLEQVDYEPDPEVRLIIVARIIMPKSAVRLSAGREKNEQMOALC
 FMAGANSIFGCKLITTPNPAEDSDMLFKKLGINREQVAKPDEITENELDRIVER
 VAARPTASDFEYDAL"
 2533. .3687
 /gene="VC1113"
 2533. .3687
 /gene="VC1113"
 /note="similar to GB:J04423 SP:P12995 PID:145426
 PID:490225 GB:000096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="8-amino-7-oxonanoate synthase"
 /protein_id="AAF94272.1"
 /db_xref="GI:9655584"
 /translation="MTLRPFSRTFOALNREHSLRRLIOLIDRSACGELLQODPYL
 NFSNDYLGIANDELVAOMQGLARFGAGSGSPVYTFSSHAHELDELCHMDYTE
 RAYLFSGFSANALLFTLLEKEDLLDQKLNHSLMEAGMSPATMKRKFINDTEHL
 ROLLHESNLVTEGVFSMDGDCAPLAQIATLVKORAMLMDADAGIGLVGDSGGS
 CQAGIHPLELVTEGKAFGLAGAVICDAHVGDYLPQAFRIHHVYSTAMPPOAHALT
 HAARMIOSSMREOLAEILACFEOCRNIPGVATQPIKPMILGSSSDSPASASHSL
 KOOGIMSAIRPPTVPGSARLKITTAHTQAOIRLSEQLKHVEMKPT"
 3549. .4487

CDS
 /gene="VC1114"
 3549. .4487
 /gene="VC1114"
 /note="similar to GB:J04423 SP:P12995 PID:145427
 PID:490221 GB:000096; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="biotin synthesis protein B10C"
 /protein_id="AAF94273.1"
 /db_xref="GI:9655585"
 /translation="MECHSTTHRACRQCAFTHYHDNGAYPSADPPLIGAVYCDGEG
 HSMATFELCELELDKSAIAQAFGRKAHAHYDHAARQVGRGLLQKMSCLGRLV
 DLGGTGYSFALLRERGAQVYADISHAMLEKORCGDEGMYDLADAEOLPFAASAC
 FDMVFSSLAQWCEDELPLSETRIVLKPFGCAFSLTLDGSLFELEDMRSDHRRH
 INQFISINQKIALAQGCSOHHLDLAATVWYEAAPMLMRDLKLGIGANHVSGRSTGL
 ISRRLAKVELAQSFRNQGGLVATYQCLGVHHR"
 4463. .5179
 /gene="VC1115"
 4463. .5179
 /gene="VC1115"
 /note="similar to GB:568059 SP:P13000 PID:490223
 PID:145428 PID:460901; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="dehydrobiotin synthetase"
 /protein_id="AAF94274.1"
 /db_xref="GI:9655586"
 /translation="MFRGSPMRNIFLAGTDTGKTVASAKIIOALATHTNIATGY
 KPAASDMEFGYGRNSDALHMKATYDMPYEDVNPVALYPTSPHIAAKRENTID
 YALLSKISKIKHQMELVYVEGAGRRPVTSSDCSLTSDGKERRPLVLVGTIKGL
 SHATLEAETRADGELVGVIANRINPGEHVAETIEHLEGRLPKLGELPYMPKAK
 KDELKFEQLDLHLEPDVA"
 complement(5261. .5470)
 /gene="VC1116"
 complement(5261. .5470)
 /gene="VC1116"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF94275.1"
 /db_xref="GI:9655587"
 /translation="MSRRLTNRKGMHAHSHYQMAIDLCHLGISEQARQQLG
 EKEQTLQOQIVDTQSHLMGLSAER"
 5697. .6560
 /gene="VC1117"
 5697. .6560
 /gene="VC1117"
 /note="similar to GB:M58470 SP:P23894 PID:146412 GB:000096
 PID:1736470; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="heat shock protein HtrX"
 /protein_id="AAF94276.1"
 /db_xref="GI:9655588"
 /translation="MKRLLPLATLAVLVISVNTVAVTWGMOGSSGLVMAA
 VYRGGATISLMKSMALRSVGVYIDTPRNEBHMLLETVKQANQAGIGFTVAL
 YDAPDMNAFATGAKRDSLVAVSTGLNNMRDEAEVLAIEVSHIANGDVITMLQ
 GVNVNTFVFLSRFTANIVASRDSGEGSNMNVFGVSMVLELFGFLASFTIMWYSR
 HREFADAGAOLVGKHKMIALERLKWGQSHLEGSMMARGITGKRSLSLMTTHPP
 LEKRTALRNM"
 6770. .7198
 /gene="VC1118"
 6770. .7198
 /gene="VC1118"
 /note="similar to PID:1653608; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="transcriptional regulator, putative"
 /protein_id="AAF94277.1"

gene

```

/db_xref="GI:9655589"
/translation="MSIMDVQVNAQFYQOLDKSQLHRIEYHPDVPEDAHRIEES
DALYQFLNRYONHCTFTTHEQYVNEGAFVWTHLRHPKAKGBOVDVKKVSHL
HFAEKVYHRYFDGEMLYEQLVLCQVIRAKIRLIGQ"
gene
7195..7920
/gene="VC1119"
7195..7920
/gene="VC1119"
/notes="similar to SP:P25970; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, short-chain
dehydrogenase/reductase family"
/protein_id="AAFP94278.1"
/db_xref="GI:9655590"
/translation="MTAVETGATSGIGKQALADYAKQGMVYIACGRNQSVLSDHTQ
YANIFLADVDVHPTKALALQPCQPELWLNAGDEYIDDKMVDYTLARVFNIN
VGVAAVIEGIDQPHLSCGRVAVGSIASELAPRAEYAGSKAAVAVATLRLQDMR
PIGEVTFTEPGEFVAPLTDRTFAMPIITVERAAQELKAGLNGSVQLVFPKRFVW
LRLGLALPYAMQGRVRLRLKA"
gene
7931..9211
/gene="VC1120"
7931..9211
/gene="VC1120"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11

```

```

Query Match          76.8%; Score 19.2; DB 1; Length 12891;
Best Local Similarity 87.5%; Pred. No. 3.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 ggcgaagcaggtcagccagcat 24
|||||
Db 1575 GCCGACGACGTCATCGTTCAT 1598

```

```

RESULT 21
HSJ324017/c          DNA          PRI          09-FEB-2001
LOCUS
DEFINITION
Human DNA sequence from clone RP3-324017 on chromosome 20 contains
the 5' end of the ID1 gene (Inhibitor of DNA binding 1, dominant
negative helix-loop-helix protein), a putative novel gene, the 3'
end of a novel gene encoding two isoforms (isoform 2 is similar to
the Drosophila CG11840) a WCT-1 (multiple copies in a T-cell
malignancies) pseudogene, 5 Cpg islands, ESTs, STS and GSSs,
complete sequence.

```

```

ACCESSION
AL110115
VERSION
AL110115.38 GI:10120310
KEYWORDS
HTG; Cpg island; ID1.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

```

```

REFERENCE
1 (bases 1 to 79419)
AUTHORS
Dunn, M.
TITLE
Direct Submision
JOURNAL
Submitted (06-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
UK
requests: clonerequest@sanger.ac.uk

```

COMMENT

On Sep 13, 2000 this sequence version replaced gi:998809.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

FEATURES

source

was generated from part of bacterial clone conlgs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HEP/Chr20>
IMPORTANT: This sequence is not the entire insert of clone
RP3-324017. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP3-324017 is at 79419 in this
sequence. The true left end of clone RP5-1093G12 is at 75615 in this
sequence. The true right end of clone RP5-1093G12 is at 100 in this
sequence. This sequence was finished as follows unless otherwise
noted: all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.
RP3-324017 is from the library RPCI-3 constructed by the group of
Pleiter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pCYRAC2.

Location/Qualifiers

repeat_region

/note="AluV repeat: matches 2. 54 of consensus"
57. 948
/note="AluX repeat: matches 1. 295 of consensus"

mRNA

join(<1082..1180,11777..11859,18545..18633,22627..22712,
22805..22930,23662..23719,28344..28427,33209..33245,
35232..35334,39808..39893,42718..42718..43163)
/gene="DJ324017.1"
/note="Continues in Em:AL121751 as DJ1093G12.7.mRNA
match: ESTs: Em:AA530502 Em:W51693 Em:NA40134 Em:AA617446
Em:AI005903 Em:AM047991 Em:AA156738 Em:AM048017
Em:AM609155"

mRNA

/evidence="not_experimental
/product="DJ324017.1.1 (A novel protein, isoform 1)"
join(<1082..1180,11777..11859,18545..18633,22627..22712,
22805..22930,23662..23719,28344..28427,33209..33245,
35232..35334,39808..39893,42718..42718..43163)
/gene="DJ324017.1"
join(<1082..1180,11777..11859,18545..18633,22627..22712,
22805..22930,23662..23719,28344..28427,33209..33245,
35232..35334,39808..39893,42718..42718..43163)
/gene="DJ324017.1"
/note="match: ESTs: Em:AI005903 Em:AA156738 Em:AI155044
Em:AM047991 Em:R67144 Em:AI1790178 Em:AA008656 Em:NA40134
Em:W51693 Em:AA617446 Em:AA019983 Em:AA269745"
/evidence="not_experimental
/product="DJ324017.1.2 (A novel protein similar to
Drosophila CG11840, isoform2)"
join(1082..1180,11777..11859,18545..18633,22627..22712,
22805..22930,23662..23719,28344..28427,33209..33245,
35232..35334,39808..39893,42718..42718..43163)
/gene="DJ324017.1"
join(<1082..1180,11777..11859,18545..18633,22627..22712,
22805..22930,23662..23719,28344..28427,33209..33245,
35232..35334,39808..39893,42718..42718..43163)
/gene="DJ324017.1"
/note="match: Proteins: Tr:O9VP07"
/codon_start=1
/evidence="not_experimental
/product="DJ324017.1.1 (A novel protein, isoform 1)"
/protein_id="CAC14948.1"
/db_xref="GI:11121545"

gene

CDS

transl_table=1
/translation="NASDPEPTTSRDARPPPIIASCTLGLVLFKIFSOEYINLL
SMFVVLILALSHITISPPMKFPASPNQYOLFLFGSGENKEEINVEPDKL
VCLGSSIVGWYLLRKHMILNLEGLAFSLNGVLEILANNVSGCILLGLFYDVF
WVEGNNWVTVAKSFEAPIKIVPDLLEKLENNPAMGLGVGVIPGIFALLRF
DISLKNHTYFYTSFAVITGICITPIMIFHHADPALLIYVACIGPFLVAAK
GEVIMFESSAEILPRLTHPTPTTSGSPASLADSMOOLAGLPPRRRPNPSAIY
EESNRKDAVAATESKEGTASASGLEKKK"

```

CDS
join(<1082..1180,1177..1185,18545..18633,22627..22712,
22805..22930,23662..23719,28344..28427,33209..33245,
35232..35334,39808..39893,42718..42817)
/gene="d3324017.1"
/note="match: proteins: Tr:O9VPQ7"
/codon_start=1
/evidence=not_experimental
/product="d3324017.1.2 (A novel protein similar to
Drosophila Cg11840, isoform2)"
/protein_id="CAC14947.1"
/db_xref="GI:11121544"
/translacion="NASDMPETISRDARPPIIASCTLLGTYLFFKIFSEYINLL
SMYPRVGLIASHSTISPPMKFFPASPNNKOTLTQSGENKEEINFEPTKDL
VCLGSSIVGWYILRKHIANNLFGAFSLNGVELLNNVSTGILLGLTYDF
WFGINWVIVAKSFPAIKLVFPQDLKELANNFMALGIDVVIPIGIFALLRF
DISLKNHTFYTSFAVIFGLITFIHMFPHADPALIYVPCAGICGPVVALAK
GEVTMFSEESNPKDPAVYESKGTASAKLEKKEK"
1915..2083
repeat_region
/note="AluYb repeat: matches 134..300 of consensus"
2084..2374
repeat_region
/note="AluSg repeat: matches 1..291 of consensus"
2375..2503
repeat_region
/note="AluYb repeat: matches 5..134 of consensus"
2593..2788
repeat_region
/note="AluYb/FRAM repeat: matches 197..308 of consensus"
2835..3142
repeat_region
/note="AluYb repeat: matches 1..309 of consensus"
3232..3366
repeat_region
/note="AluSx repeat: matches 1..122 of consensus"
3367..3595
repeat_region
/note="AluSc repeat: matches 83..306 of consensus"
3596..3689
repeat_region
/note="AluSx repeat: matches 122..219 of consensus"
3690..3721
repeat_region
/note="Alu repeat: matches 90..115 of consensus"
3784..3826
repeat_region
/note="Alu repeat: matches 251..293 of consensus"
4002..4306
repeat_region
/note="AluSc repeat: matches 1..303 of consensus"
4439..4634
repeat_region
/note="L1MC/D repeat: matches 5453..5636 of consensus"
4646..4936
repeat_region
/note="AluSx repeat: matches 1..297 of consensus"
4997..5158
repeat_region
/note="FRAM repeat: matches 1..162 of consensus"
5188..5319
repeat_region
/note="AluSg/x repeat: matches 171..308 of consensus"
5360..5570
repeat_region
/note="AluYb repeat: matches 84..298 of consensus"
5743..6042
repeat_region
/note="AluSc repeat: matches 1..300 of consensus"
6791..6818
repeat_region
/note="MER88 repeat: matches 433..459 of consensus"
6819..7124
repeat_region
/note="AluSx repeat: matches 1..304 of consensus"
7125..7393
repeat_region
/note="MER88 repeat: matches 75..433 of consensus"
7444..7475
repeat_region
/note="MER88 repeat: matches 1..31 of consensus"
7476..7775
repeat_region
/note="AluYb repeat: matches 1..302 of consensus"
7776..8099
repeat_region
/note="MER88 repeat: matches 31..353 of consensus"
8100..8399
repeat_region
/note="AluYb repeat: matches 1..302 of consensus"
8402..8704
repeat_region
/note="AluY repeat: matches 1..302 of consensus"
8705..8789
repeat_region
/note="MER88 repeat: matches 353..447 of consensus"
8884..8959
repeat_region
/note="38 copies 2 mer aa 64% conserved"
9286..9467
repeat_region
/note="L2 repeat: matches 2559..2750 of consensus"

```

```

repeat_region 9679..9961
/note="AluSg repeat: matches 1..283 of consensus"
9985..10643
repeat_region
/note="L2 repeat: matches 1971..2748 of consensus"
10684..10999
repeat_region
/note="L2 repeat: matches 2146..2461 of consensus"
11951..12071
repeat_region
/note="MIR repeat: matches 36..147 of consensus"
12257..12425
repeat_region
/note="MIR repeat: matches 8..175 of consensus"
12483..12679
repeat_region
/note="MIR repeat: matches 3..205 of consensus"
13120..13406
repeat_region
/note="AluSc repeat: matches 9..294 of consensus"
13446..13669
repeat_region
/note="MIR repeat: matches 2..238 of consensus"
13796..14418
misc_feature
/note="match: GSS: Em:A0749725"
14311..14598
repeat_region
/note="AluSx repeat: matches 5..303 of consensus"
14799..15110

```

```

Query Match 76.8%; Score 19.2; DB 93; Length 79419;
Best Local Similarity 87.5%; Pred. No. 2.6e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 1 ggcgacgagtgatcgccagcat 24
Db 16785 GCGGACGAGTGCGACGCGCAT 16762

```

```

RESULT 22
AP003003
LOCUS
DEFINITION
AP003003 349116 bp DNA BCT 03-FEB-2001
Mesorhizobium loti DNA, complete genome, section 10/21, complete
sequence.
ACCESSION
AP003003 BA000012
VERSION
AP003003.1 GI:11994978
KEYWORDS
HTG.
SOURCE
Mesorhizobium loti (strain:MAFP303099) DNA.
Mesorhizobium loti
ORGANISM
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE
1 (sites)
Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpoto,S., Sugimoto,M.,
Takeuchi,C., Yamada,M. and Tabata,S.
Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
DNA Res. 7 (6), 331-338 (2000)
2 (bases 1 to 349116)
JOURNAL
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp,
URL:http://www.kazusa.or.jp/rhizobase/, Tel:81-438-52-3935,
Fax:81-438-52-3934)
TITLE
Direct Submission
AUTHORS
Kaneko,T.
JOURNAL
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp,
URL:http://www.kazusa.or.jp/rhizobase/, Tel:81-438-52-3935,
Fax:81-438-52-3934)
FEATURES
source
1..349116
Location/Qualifiers
/organism="Mesorhizobium loti"
/strain="MAFP303099"
/db_xref="taxon:381"
BASE COUNT 63710 a 112865 c 109235 g 63306 t
ORIGIN
Query Match 76.8%; Score 19.2; DB 2; Length 349116;
Best Local Similarity 87.5%; Pred. No. 2.1e+02;

```

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ggcgcagcagtcagccagcat 24
 |||
 Db 213715 GCGCAGCAGTTCATCGCTCCAT 213738

RESULT 23
 PSEHRIA/C 3603 bp DNA BCT 21-OCT-1993
 DEFINITION Pseudomonas syringae hprp1 genes, complete cds.
 ACCESSION L11582
 VERSION L11582.1 GI:151267
 KEYWORDS hprp1 gene; protein translocation.
 SOURCE Pseudomonas syringae (strain 61) DNA.
 ORGANISM Pseudomonas syringae
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.

REFERENCE 1 (bases 1 to 3603)
 AUTHORS Huang, H.-C., Xiao, Y., Lin, R.-H., Lu, Y., Hutcheson, S.W. and
 Collmer, A.
 TITLE Characterization of the Pseudomonas syringae pv. syringae 61 hprp1
 and hprp2 genes: homology of Hprp1 to a superfamily of proteins
 associated with protein translocation
 JOURNAL Mol. Plant Microb. Interact. 6, 515-520 (1993)
 FEATURES
 source Location/Qualifiers
 1..3603
 /organism="Pseudomonas syringae"
 /db_xref="taxon:317"
 1330..3417
 /gene="hprp1"
 1330..3417
 /note="hprp1"
 /note="hyper-sensitive response and pathogenicity gene,
 membrane-spanning protein; hprp1 superfamily proteins are
 associated with protein translocation; putative"
 /codon_start=1
 /transl_table=11
 /product="hprp1"
 /protein_id="AA03015.1"
 /db_xref="GI:151268"
 /translation="MNRVINFNMVALSAMRSELVAFVIAIVFMITPILGLID
 VLAIVNICISCLILAMHLPPLAFSTEPVALITLTERIALSVSTRLLINDAG
 HIVEAFGQVVGNGLVAGLVITLTVNPLVITKGSFRAVEVGRFLDMPCKOS
 IDSILRANLITVHEARKRAELNKSQIFGAMDGMKFNVDATSLITVAINMGCI
 STGVLDHNNAGDALQITVLTIGGLIQAIPALISVSGMITVRVNTAGVEANI
 GEIAEQITISQKAMIVASVAMLGFAALPGMTGVITITAIICGGGLLQKAPKA
 DEQRTAAVAPEMNGKEDRTESPSQVLFQHPGDSQAIQALVSEIKRRNRLVQY
 GLTLPSTIEHVDIAPDEFRTYDVMLKATFQSHVAVAROLEGENPAALPGN
 TDRODQWMLPABOSGELNPSSTLTIERMERALQSCAPQFISGLOETKAILSMLE
 EDPRLAQMQRVLTITRESAVLQRLASCVCPLRAIRVATLTIEMCOHERPNVITDY
 VIALKSOIYHOYCGABGLQVLTPEBGLRDLGRLROTETFPALSNESOMLVQO
 LHTAFVRAPEQAVLIVADLRSPRLTLIRREYHVPVLSFAELISNAKVRVMGREDL
 EDLLEPLNEHHA"

BASE COUNT 772 a 1093 c 1035 g 703 t

Query Match 76.0%; Score 19; DB 3; Length 3603;
 Best Local Similarity 100.0%; Pred. No. 4.9e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ggcgcagcagtcagccagcat 24
 |||
 Db 1522 GCGAGTGCATCGCCAGCAT 1504

RESULT 24
 MTCY180/c 44201 bp DNA BCT 17-JUN-1998
 LOCUS MTCY180
 DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 85/162.
 ACCESSION Z97193 AL133456
 VERSION Z97193.1 GI:3261816

KEYWORDS
 SOURCE
 ORGANISM
 Mycobacterium tuberculosis.
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
 Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1 (bases 1 to 44201)
 AUTHORS Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
 Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
 Tekait, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
 Connor, R., Davies, R., Devlin, K., Feldwell, T., Gentles, S.,
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
 Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
 Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
 Barrell, B.G.
 TITLE Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence
 JOURNAL Nature 393 (6685), 537-544 (1998)
 MEDLINE 98295987
 REMARK Erratum: [[published erratum appears in Nature 1998 Nov
 12;396(6707):190]]
 REFERENCE 2 (bases 1 to 44201)
 AUTHORS Parkhill, J.
 TITLE Direct Submission
 JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
 tuberculosis sequencing and mapping teams: Sanger Centre, Wellcome
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced g1:2225942.

NOTES:
 Details of M. tuberculosis sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
 been renumbered from the original cosmid subclonings but the old
 gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of TB genes
 implemented in TBparse (Krogh) supplemented with visual inspection
 of positional base preference in codons, especially where there is
 an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct
 initiation codon. Where possible we choose an initiation codon
 (atg, gtg, or ttg) which is preceded by an upstream ribosome
 binding site sequence (optimally 5-13bp before the initiation
 codon). If this cannot be identified we choose the most upstream
 initiation codon.

FEATURES
 source Location/Qualifiers
 1..44201
 /organism="Mycobacterium tuberculosis"
 /strain="H37Rv"
 /db_xref="taxon:1773"
 <1..42142
 /organism="Mycobacterium tuberculosis"
 /strain="H37Rv"
 /db_xref="taxon:1773"
 /clone="Y180"
 23..460
 /gene="Rv1873"
 23..460
 /note="Rv1873, (MTCY180.45c), len: 145. Unknown, TBparse
 score is 0.909"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv1873"
 /protein_id="CAB10052.1"
 /db_xref="GI:2225987"
 /db_xref="SPTREMBL:O07756"
 /translation="MKSASDPDLCKRKYAOAPYRSVPELRAGRGHMMFVFPQ
 LRGSSFLAVRTISSLEDAVTLQDLDLGPRIHETGVLVNOYQGSIEIFGPPDD
 LKLCSSMTLIFARITANDQVALLAKYGGGDEDRVALLAVT"
 517..520

RBS

```

gene
  /note="possible RBS, GGAG, for Rv1874"
  /gene="Rv1874"
  533..1219
  /db_xref="SPTREMBL:007753"
CDS
  /gene="Rv1874"
  /note="Rv1874, (MTCY180.44c), len: 228. Unknown, Tbpase score is 0.928"
  /codon_start=1
  /transl_table=11
  /product="Hypothetical protein Rv1874"
  /protein_id="CAB10067.1"
  /db_xref="GI:2225986"
  /db_xref="SPTREMBL:007755"
  /translation="MLMRPDDDDMCARQARQAVADALLGLGACISINVDSTYRSL
  MTLTLVPPVAAYVSLMTQCCGQVDAALRLAOCEDCLGAVLVTSVPLTPPSLVE
  SGRTPGLANIALRRPDGLDQATWLIRWQDHQVALAEATGTYQNMVVRALTPR
  APGAGIVLEELPPVAATTLDAKAFGADNDLRNRISRMVASTSAFANONIDIVPTS
  RYVETPFKX"
  1218..1221
  /note="possible RBS, GAGG, for Rv1875"
  1230..1673
  /gene="Rv1875"
  1230..1673
  /note="Rv1875, (MTCY180.43c), len: 147. Some similarity
  to Z55841MTC165.22 (147 aa) opt: 178 z-score: 240.3 E():
  7.4e-06; 26.9% identity in 130 aa overlap. Tbpase
  score is 0.908"
  /codon_start=1
  /transl_table=11
  /product="Hypothetical protein Rv1875"
  /protein_id="CAB10051.1"
  /db_xref="GI:2225985"
  /db_xref="SPTREMBL:007754"
  /translation="MTLNFAALAAERGLAVSVTRADQTVQASLVNGLLPHPV
  GEPISGFTYKQVGLNLRARPOLAVTFRNGWQAVTEGRAQLVGPDDPPWLVDER
  LRLLEVFTTAAAGTTHDDMDYDRVMAQEQRAVALLTPRTIRYNSG"
  2179..2183
  /note="possible RBS, AGGAG, for Rv1876"
  2189..2668
  /gene="bfrA"
  2189..2668
  /gene="bfrA"
  /note="Rv1876, (MTCY180.42c), bfrA, len: 159. Function:
  bacterioferritin. FASTA results: BFR_MYCLE P43315
  bacterioferritin (bfr) (159 aa) opt: 958 E(): 0; 90.6%
  identity in 159 aa overlap. Tbpase score is 0.913"
  /codon_start=1
  /transl_table=11
  /product="bfrA"
  /protein_id="CAB10050.1"
  /db_xref="GI:2225984"
  /db_xref="SPTREMBL:008465"
  /translation="MOGDPIVRLNLNOLSETLAINOYFLHKKMODNMNGFELAH
  RASFDPMRAEEETRIILDLGILPNVORSGTARICOTLRDEPDLALAEVYINLTK
  PGIVMCREKODTSAVLEKIVADEBEHDYLETQLEMDKAEELYSACVSRPT"
  2742..2746
  /note="possible RBS, GGAGG, for Rv1877"
  2753..4816
  /gene="Rv1877"
  2753..4816
  /gene="Rv1877"
  /note="Rv1877, (MTCY180.41c), len: 687. Integral membrane
  protein similar to many antibiotic and drug efflux
  proteins. M. tuberculosis relatives include: MTCY3G12.01,
  MTCY98.0002c and MTCY369.27c. Contains P500217 Sugar
  transport proteins signature 2 (P500217). FASTA results:
  Q56175 (557 aa) opt: 895; E(): 0; 34.7% identity in 528
  aa overlap. Tbpase score is 0.916"
  /codon_start=1
  /transl_table=11
  /product="Hypothetical protein Rv1877"
  /protein_id="CAB10049.1"

```

```

misc_feature
  /db_xref="GI:2225983"
  /translation="MAGPTAPPTAATRAAGPLSPVRNITTFALVPGVVAATG
  TIVPALPRIVAFEGSTVQDSNAVTSYLLGGTYVVAARLGLDLGRNRVLLGVV
  VSGSLGSLQTMFLAISRALGVAGAI SVTALAAEAVPLRGRXQGLGAVF
  GVNTVTPPLIGWLTIDYLSWRAMFWINPVSLAVLVAALAVPALARPAPVLDYGI
  LVIAVATLALIMATSWGGTTYAMGSAITVGLVIGVDAALGFFVWLEGRRAALILPRL
  FGSPPVAVCCSVSEVFGAMLGATFEPYILGVDAASATAGSLRTLPVIGLLAST
  GTGLVGRGRYKIFPPVAGMALMAVAFLMSOMDEWTPPLQSLVLYVIGAGTISMO
  VLYLVONTSSPEDLGATSGVTFEPVAGSCTAFGLFPNPLDRRLGSLTSGAV
  PVPVPSPAVHLQDPSMAPIVRAAESLTQVLCANSVTVGFTLLALILEVPLTD
  IHDDADLDGFGVPRAESPEVLEIAVRMLPNVGRADATVQGGCGVAMEALL
  RITYQYRLEAVALLDIGHLVHPYQVPEVDRIVTGTVAARDIDITLTPSGRQV
  DSLAVLIRQMLDLHLAVAPGLRKOPDHOPEALQHVTDVAVLQDWEYEDIGLSSRO
  LAAT"
  3122..3199
  /gene="Rv1877"
  /note="P500217 Sugar transport proteins signature 2"
  4871..6223
  /gene="glnA3"
  4871..6223
  /gene="glnA3"
  /note="Rv1878, (MTCY180.40c), len: 450. glnA3, similar to
  many glutamine synthetases. FASTA results: GlnA_BACE
  P19064 (EC 6.3.1.2) (443 aa) opt: 497; E(): 5.2e-23; 29.0%
  identity in 331 aa overlap. Tbpase score is 0.925. Also
  similar to C-terminus of FLUG_EMBNI P38094 flag protein.
  emeriellandulians (865 aa), fasta scores: opt: 227
  z-score: 357.9 E(): 6.4e-13 (29.9% identity in 394 aa
  overlap) (note that the downstream ORF MTCY180.39c is
  similar to the N-terminus) and to several other M.
  tuberculosis proteins eg. MTCY427.03c, MTCY427.01 and
  MTCY190.31"
  /codon_start=1
  /transl_table=11
  /product="glnA3"
  /protein_id="CAB10048.1"
  /db_xref="GI:2225982"
  /db_xref="SPTREMBL:007752"
  /translation="MTATPPLAAALIAOLEABGVDTVIGTVNPAQLQAKTVPIRKN
  TPANGLASPVMTFCIDQCSIAFTADISVAGDRLRIDLSALRIDGLTAMAPAG
  FEODGTVPVACSRGTLSRIEALADAGIDAVTAGEVEFLIADADGGRIPSTLMAYG
  AGVLHEAEFVRVNAATPAAGIAIROPFREGANOFETSLADPPVAAADVLITLI
  IGRTRRHGLRVSRLSAPFAGSISGAHOFSLMSREGLPSSGTGAAGMAGAEAV
  AGVINGLDAGGIIIGSTVSGLRMPGCMGATVCMKGENEAARVPRVKGAGATG
  NNEVYVNPDSNAPYLAASAILGLAIDKRTAAVLPSETTVDPQLSDVDRRAILRL
  AADQADALVADLSKRLKLTIGDPVDAVVAVRQLEHRYGDDLPADLAKFRMAWV
  "
  6213..6216
  /gene="glnA3"
  6213..6216
  /gene="glnA3"
  Query Match
  Best Local Similarity 76.0%; Score 19; DB 3; Length 44201;
  Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
  QY 2 cgcagcaggtgcacgcga 20
  DB 25769 CCGACGACGTGCATCGCCA 25751
  RESULT 25
  AEO01827/c
  LOCUS 45704 bp DNA circular BCT 10-JAN-2001
  DEFINITION Deinococcus radiodurans R1 plasmid Cpl, complete plasmid sequence.
  ACCESSION AEO01827
  VERSION AEO01827.1 GI:6460959
  KEYWORDS
  SOURCE
  ORGANISM
  Deinococcus radiodurans.
  Deinococcus radiodurans
  Bacteria; Thermus/deinococcus group; Deinococcales; Deinococcus.
  1 (bases 1 to 45704)
  White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
  Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,

```


TITLE	Mofatt, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathavan, J., J., Lam, P., McDonald, L., Utterback, T., Zaleski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Fraser, C.M.
JOURNAL	Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1
MEDLINE	Science 286 (5444), 1571-1577 (1999)
PUBMED	20036896
REFERENCE	10567266
AUTHORS	2 (bases 1 to 45704)
TITLE	White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D., Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L., Mofatt, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M., Vamathavan, J., J., Lam, P., McDonald, L., Utterback, T., Zaleski, C., Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W., Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S., Smith, H.O., Venter, J.C. and Fraser, C.M.
JOURNAL	Direct Submission
Medical Center Dr. Rockville, MD 20850, USA	
Location/Qualifiers	
source	1. 45704
organism	"Deinococcus radiodurans"
strain	"R1"
db_xref	"taxon:1299"
plasmid	"CPI"
gene	73. .384
gene	/gene="DRC0001"
CDS	73. .384
gene	/note="identified by sequence similarity; putative"
product	"cytochrome P450-related protein"
protein_id	"AAFI2681.1"
db_xref	"GI:6460978"
translation	"MCSPPRRYSGSTDAEILPYVVEARESSPVYHLDISQNRDESEV TQAQPCPRKNVROHLSPFGKMGACLGSLVYTCRWMAHLELLSAPAGKIDVYS Q"
gene	725. .1951
gene	/gene="DRC0002"
CDS	725. .1951
gene	/note="identified by sequence similarity; putative"
product	"phosphoenolpyruvate synthase-related protein"
protein_id	"AAFI2663.1"
db_xref	"GI:6460960"
translation	"MSHLVRLDISITGIPDELGRLLQETIELCGYAMDVHRLTTP EVEAFYMWCRINAGVTNELHELFIDGDFSLRQARGFWKLHASSLRECEVEDS DNATQPOFLKROECOVLSALPIIDKREPFDLYSAFSSHLREDHEVMECPALLR RKEQKATGRLHALGLIEDPLDCAVFTLTETLOPAGISIALQALRELPARKSESDLK FKPRPRIRAPSSDDVSKLLKMGKISGTCYSVCYSVGEAFELKISMGDIIVCP ETGPDWTPLESTAGGLVLESPFAGMLSHALVAREKRIPIALICISACSLDGRVYL DETGLVEIGRFEEM"
gene	1938. .2801
gene	/gene="DRC0003"
CDS	1938. .2801
gene	/note="identified by sequence similarity; putative"
product	"phosphoenolpyruvate synthase-related protein"
protein_id	"AAFI2680.1"
db_xref	"GI:6460977"
translation	"MNQCDCKNSLLDCSCIPSLIPLFLVDRMTLEERVGKAANGE LAGMGFEIPNGCVITIDVFDALNVAISLPVGKRGKGFSTMANLHPDQILPHLEMAER LGPVAVRSSAVFEDSDASVAGOLSSILNVEGFPDITVAVECMHSJEGGVKTKYSK LHNDGIDRLMAVIVOEQVPRKAGVFAHPTGPEHTVIEAVSGIKNLKVDVGT PNHWVIDRNSREVIESHIFNYOVSEALKEALHLDGIKQVTLFGKPDIEMLKDGKP
gene	3325. .4046
gene	/gene="DRC0004"
CDS	3325. .4046
gene	/note="This region contains an authentic frame shift and is not the result of a sequencing artifact; similar to PID:1737165 percent identity: 49.68; identified by sequence similarity; putative; transposase, putative, authentic frameshift"
product	"transposase"
protein_id	"AAFI2664.1"
db_xref	"GI:6460961"
translation	"MTRGQRYIRVSTPDONTAROLDGVLEDRIFEDKASGDARRP RIQELLAYVREGDPTVIVHSMRDLARNVDLRIIVTELSGRGVREFORGVPTSEDN PMTLLITMIGNAVEPIRAONLENORGTAKAKQCKYKGRKALIPQVEQLARAYA AGEKTVLARELGVNRTLYEALKS"
gene	complement(5471. .6205)
gene	/gene="DRC0006"
CDS	complement(5471. .6205)
gene	/note="identified by sequence similarity; putative"
product	"conserved hypothetical protein"
protein_id	"AAFI2677.1"
db_xref	"GI:6460974"
translation	"MIDPHYHLDYPPVAVACEROLTYLVTTTPAMKGTLLA AAGRPHWIALGHPPEVSEERADLPFDRLTETPRVGVGDGSPSLRGTYTQQA VFOHILRCEHDGGRILSHSRKESVLCLEPNRSGTPIIHWVSGVTELRATIS LGMFVSQPTMVTQKGAALIRSMRDRVLTETDGPLDGOAALPMVDKSVYEGLS KICWIPASEVERIVKENVSRLGTVR"
gene	complement(6202. .7473)
gene	/gene="DRC0007"
CDS	complement(6202. .7473)
gene	/note="identified by glimmer2; putative"
product	"hypothetical protein"
protein_id	"AAFI2682.1"
db_xref	"GI:6460979"
translation	"MRHFLCTPNPAAVNSGSTIVALYGRRLNTHOVGASLREAVI RKMPFPSQWDFSLTSLVLAADSSPSESPPGWTRELSTISYDADIMNQAARA LEALAVYLTTRGWSITPSPGGYOYQKPEIGLIEDCVALLSGDLSLVMIDLVEG HNPLAVSGLVAGDAKORETATDGRTHIOLNNAACPTPGEDSORSRTLLFLAVY LAATATTRFOGARAVPIYINNGTALNPLTTPRISLSITRAHPYEQLOOLAVYOA VGLVVALINPALTKGEMLTECANOTVLSRWASATSCGTRFRFGIKGKGCICPOY BRAAFILMGQDREKVEVYPLGQNDHSHALFEDVRSIALAIIVVEGSGVQRFGLASIS DRIIGDRRAMAMVKRSMEELGALHRLKVP"
gene	complement(7470. .8057)
gene	/gene="DRC0008"
CDS	complement(7470. .8057)
gene	/note="identified by glimmer2; putative"
product	"hypothetical protein"
protein_id	"AAFI2692.1"
db_xref	"GI:6460989"
translation	"MSSYVKGSGHSGKTAAGRMASATAGRLFSPLSAPVGTGPGQ LRSVDPLQPSRASDITDAVIEGVQVNGTLDASRHA INXALSTLLQHPDADLL ALTADQELVELVVAODVFPHIELDGOALRQOGVSAALMLTDLRQARBYVQATVSG AFDELRRGQAITAPVQAVTORALELMEFAFG"
gene	complement(8308. .10218)
gene	/gene="DRC0009"
CDS	complement(8308. .10218)
gene	/gene="DRC0009"

```

/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAFI2683.1"
/db_xref="GI:6460980"
/translation="MADTETDRYLNTSVANVAVELIVSAGNPVSIQVAGAVG
KSSMTKLIRNLNERQSGIPASNGVENSANARSGTATPKVFEVFNMLVQGDAR
AALMDVILARELTAEEAEKQTCMDHVKDVSRLNMRGAVAHILGAAGLPGVGLIG
ETASAISGLRDKIEGDIKGAEKVAGQAVITLGLKAAQSPPOIOARSEFT
ALEKIDVVLVLDLDRCLPETTISTLEALITFLKRTAFVIAADNMIKHVRKH
PEGMDNAVINYPDKLIQVPRVPLSTQVRAVILLIVEDSELEKRDVAVATY
NKALREGSGTRLDGDFINOHPLPELKPRLIADILAPLATANGIDSNPRIK
EEMETALAGEPLSLPSPWNEEFVKMLALPVLSDIPIVMSLIRQVGLDPNGKPSFLQ
TKAASDVILAMKANGPSGQLRQELALSPROGVMDTLIREAQREERMGAPRILDD
CLAVSRLLHDSQAKLAPLKNKPMOTLPSPILPKISGEPMMAELFTDMKISGARRHY
NAIEAERRRRRA"
complement(10441..10998)
/gene="DRC0010"
complement(10441..10998)
/gene="DRC0010"
/note="identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAFI2665.1"
/db_xref="GI:6460962"
/translation="MRGLGGLGAYILFYGPHELYIPVTANQONTPIYIGKAV
FKGDRITGGAKKAVLENSLYKRLYHRSIAVENLDVADYFVVFVPTTLISAVES
VLISKFVPMNKHIDGFCNHDGSGRGSRNOKRSVMDVYHGRSMATRMSNLAIYDAEL
RSRISARIHAKVAVKAIIEASEK"
complement(11512..12403)
/gene="DRC0011"
/note="this region contains a pseudogene, one or more
premature stops, and is not the result of a sequencing
artifact; similar to PID:1162996 percent identity: 52.81;
identified by sequence similarity; putative;transposase,
pseudogene"
12794..13333
/gene="DRC0012"
12794..13333
CDS
/gene="DRC0012"

Query Match
Best Local Similarity 75.2%; Score 18.8; DB 1; Length 4570;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 gcagcaggtgcagccagcatt 24
||||| ||||| ||||| |||||
Db 19924 GCAGCAGTTGCAGCCGACGACAT 19903

gene
12794..13333
/gene="DRC0012"
12794..13333
CDS
/gene="DRC0012"

Query Match
Best Local Similarity 75.2%; Score 18.8; DB 1; Length 110000;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 cagcaggtgcagccagcatt 25
||||| ||||| ||||| |||||

```

```

Db 22665 CAGCAGAACATCCGACGACATT 22686

RESULT 27
AC025404
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

```

```

AC025404 148900 bp DNA HTG 24-MAR-2000
Homo sapiens chromosome 1 clone RP11-301H2 map 1, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
AC025404
AC025404.2 GI:7321650
HTG; HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Unpublished
2 (bases 1 to 148900)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bede, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
Dodgson, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehotzky, J.,
Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
Meidrim, J., Meneses, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,
Murphy, T., Naylor, T.M., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Tamas, J.,
Testfaye, S., Theodore, J., Tittel, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 24, 2000 this sequence version replaced gi:7210127.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 301_H2
Center clone name: 301_H2
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Assembly: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 13985 bases at least Q40
Consensus quality: 144213 bases at least Q30
Consensus quality: 145912 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 147300; sum-of-ctnigs
Quality coverage: 4.3 in Q20 bases; agarose-fp
Quality coverage: 4.4 in Q20 bases; sum-of-ctnigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1610: contig of 1610 bp in length
* 1611 1710: gap of 100 bp
* 1711 4083: contig of 2373 bp in length
* 4084 4183: gap of 100 bp
* 4184 6794: contig of 2611 bp in length
* 6795 6894: gap of 100 bp
* 6895 10293: contig of 3399 bp in length
* 10294 10393: gap of 100 bp
* 10394 14279: contig of 3886 bp in length
* 14280 14379: gap of 100 bp
* 14380 17741: contig of 3362 bp in length
* 17742 17841: gap of 100 bp
* 17842 21194: contig of 3353 bp in length
* 21195 21294: gap of 100 bp
* 21295 27169: contig of 5875 bp in length
* 27170 27269: gap of 100 bp
* 27270 34236: contig of 6967 bp in length
* 34237 34336: gap of 100 bp
* 34337 42772: contig of 8436 bp in length
* 42773 42872: gap of 100 bp
* 42873 52351: contig of 9479 bp in length
* 52352 52451: gap of 100 bp
* 52452 63753: contig of 11302 bp in length
* 63754 63853: gap of 100 bp
* 63854 74258: contig of 10405 bp in length
* 74259 74358: gap of 100 bp
* 74359 87809: contig of 13451 bp in length
* 87810 87909: gap of 100 bp
* 87910 102652: contig of 14743 bp in length
* 102653 102752: gap of 100 bp
* 102753 119718: contig of 16966 bp in length
* 119719 119818: gap of 100 bp
* 119819 148900: contig of 29082 bp in length.

```

FEATURES

```

source
1. 148900
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="1"
  /map="1"
  /clone_1lb="RP11-301H2"
  /clone_1lb="RP11-11 Human Male BAC"
  1. 1610
  /note="assembly_fragment"
  misc_feature
  1711. 4083
  /note="assembly_fragment"
  misc_feature
  4184. 6794
  /note="assembly_fragment"
  misc_feature
  6895. 10293
  /note="assembly_fragment"
  misc_feature
  10394. 14279
  /note="assembly_fragment"
  misc_feature
  14380. 17741
  /note="assembly_fragment"
  misc_feature
  17842. 21194
  /note="assembly_fragment"
  misc_feature
  21295. 27169
  /note="assembly_fragment"
  misc_feature
  27270. 34236
  /note="assembly_fragment"
  misc_feature
  34337. 42772
  /note="assembly_fragment"
  misc_feature
  42873. 52351
  /note="assembly_fragment"
  misc_feature
  52452. 63753
  /note="assembly_fragment"
  misc_feature
  63854. 74258
  /note="assembly_fragment"
  misc_feature
  74359. 87809
  /note="assembly_fragment"
  misc_feature
  87910. 102652

```

```

/note="assembly_fragment
clone_end:SP6
vector_side:right"
102753. 119718
/note="assembly_fragment
clone_end:r7
vector_side:left"
119819. 148900
/note="assembly_fragment"
BASE COUNT 43768 a 28628 c 29209 g 45693 t 1602 others
ORIGIN

```

```

Query Match 75.2% Score 18.8; DB 69; Length 148900;
Best Local Similarity 90.9%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy 4 cagcaggtgacgcagcagcatt 25
||||| |||||||||
Db 74753 CAGCAGGACATCGCAGCAT 74774

```

```

RESULT 28
AL365213 176687 bp DNA HTG 23-JAN-2001
LOCUS Homo sapiens chromosome 1 clone RP11-224019, *** SEQUENCING IN
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

```

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

```

COMMENT

```

Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba224019
----- Summary Statistics
Sequencing vector: XGAP4; version 4.5
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 167635 bases at least Q40
Consensus quality: 170858 bases at least Q30
Consensus quality: 172773 bases at least Q20
Insert size: 174987; sum-of-contigs
Insert size: 160662; 29.8% error; agarose-IP
Quality coverage: 3.77x in Q20 bases; sum-of-contigs Quality
coverage: 4.31x in Q20 bases; agarose-IP
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 19212: contig of 19212 bp in length
* 19213 19312: gap of 100 bp
* 19313 36455: contig of 17143 bp in length
* 36456 36555: gap of 100 bp
* 36556 41590: contig of 5035 bp in length

```

```

* 41591 41690: gap of 100 bp
* 41691 47894: contig of 6204 bp in length
* 47895 47994: gap of 100 bp
* 47995 50477: contig of 2483 bp in length
* 50478 50577: gap of 100 bp
* 50578 57800: contig of 7223 bp in length
* 57801 57900: gap of 100 bp
* 57901 71528: contig of 13628 bp in length
* 71529 71628: gap of 100 bp
* 71629 76173: contig of 4545 bp in length
* 76174 76273: gap of 100 bp
* 76274 105540: contig of 29267 bp in length
* 105541 105640: gap of 100 bp
* 105641 111894: contig of 6254 bp in length
* 111895 111994: gap of 100 bp
* 111995 123775: contig of 11781 bp in length
* 123776 123875: gap of 100 bp
* 123876 132910: contig of 9035 bp in length
* 132911 133010: gap of 100 bp
* 133011 136379: contig of 3369 bp in length
* 136380 136479: gap of 100 bp
* 136480 144993: contig of 8514 bp in length
* 144994 145093: gap of 100 bp
* 145094 153649: contig of 8556 bp in length
* 153650 153749: gap of 100 bp
* 153750 157198: contig of 3449 bp in length
* 157199 157298: gap of 100 bp
* 157299 161364: contig of 4066 bp in length
* 161365 161464: gap of 100 bp
* 161465 176687: contig of 15223 bp in length.

```

FEATURES

```

Source
Location/Qualifiers
1..176687
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone_lib="RPC1-11.1"
1..19212
/note="assembly_fragment:01465"
fragment_chain:1"
19313..36455
/note="assembly_fragment:01127"
fragment_chain:1"
36556..41590
/note="assembly_fragment:00159"
fragment_chain:1"
41691..47894
/note="assembly_fragment:00227"
fragment_chain:1"
47995..50477
/note="assembly_fragment:00174"
fragment_chain:1"
50578..57800
/note="assembly_fragment:01341"
fragment_chain:1"
57901..71528
/note="assembly_fragment:00452"
fragment_chain:1"
71629..76173
/note="assembly_fragment:01981"
fragment_chain:2"
76274..105540
/note="assembly_fragment:00319"
fragment_chain:2"
105641..111894
/note="assembly_fragment:01486"
fragment_chain:2"
111995..123775
/note="assembly_fragment:00713"
fragment_chain:3"
123876..132910
/note="assembly_fragment:01542"
fragment_chain:3"

```

```

misc_feature 133011..136379
/note="assembly_fragment:00535"
136480..144993
/note="assembly_fragment:00623"
145094..153649
/note="assembly_fragment:01468"
153750..157198
/note="assembly_fragment:01546"
157299..161364
/note="assembly_fragment:02010"
161465..176687
/note="assembly_fragment:00187"
clone_end:"7"
vector_side:right"
BASE COUNT 52919 a 35491 c 34728 g 51832 t 1717 others
ORIGIN

```

```

Query Match 75.2% Score 18.8; DB 80; Length 176687;
Best Local Similarity 90.9%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 4 cagcagtgatcgccagcatt 25
||||| |||||||||
Db 58230 CAGCAGACATCGCCAGCATT 58251

```

```

RESULT 29
AF250770 965 bp DNA BCT 31-JAN-2001
LOCUS AF250770
DEFINITION Bacterium pCOSH1 DAPA-aminotransferase (bica) and
biotin synthase (bicaB) genes, partial cds.
ACCESSION AF250770
VERSION AF250770.1 GI:12620104
KEYWORDS uncultured bacterium pCOSH1.
Bacteria: environmental samples.
ORGANISM uncultured bacterium pCOSH1.
SOURCE uncultured bacterium pCOSH1.
TITLES 1 (bases 1 to 965)
REFERENCE Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Strelt,W.R.
Direct cloning from enrichment cultures, a reliable strategy for
isolation of complete operons and genes from microbial consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE 20575196
PUBMED 11133432
REFERENCE 2 (bases 1 to 965)
AUTHORS Entcheva,P., Liebl,W. and Strelt,W.R.
TITLES Direct Submission
JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
Source
Location/Qualifiers
1..965
/organism="uncultured bacterium pCOSH1"
/db_xref="taxon:143796"
/clone="pCOSH1"
/note="unknown organism, cosmid clone derived from
environmental consortium"
complement(<1..438)
/gene="bica"
complement(<1..438)
/gene="bica"
/note="7,8-diaminopelargonic acid
synthetase-aminotransferase, similar to bica"
/codon_start=1
/transl_table=1
/product="DAPA-aminotransferase"
/protein_id="AAG60563.1"
/db_xref="GI:12620105"
/transl_table="MTODDLAFQDQNHIMPTSTGRXYRVVSAAXHOCRRRLVD
GMSMMAIHGXNRPRLNAAMKADIDMSNMREGTTHQRAVLCRKHVAMTPDALSC
VFADSGSVAKEVADENKASTAPLRADRPERRHNAFFPIKIQ"
525..>965
/gene="bicaB"

```

gene

VERSION M60166.1 GI:170463
 KEYWORDS H+-ATPase; plasma membrane protein; transmembrane protein.
 SOURCE L.esculentum, cDNA to mRNA.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 REFERENCE 1 (bases 1 to 3229)
 AUTHORS Ewing,N.N., Wimmers,L.E., Meyer,D.J., Chetelat,R.T. and
 Bennett,A.B.
 TITLE Molecular cloning of tomato plasma membrane H+-ATPase
 JOURNAL Plant Physiol. 94, 1874-1881 (1990)
 FEATURES Location/Qualifiers
 source 1..3229
 /organism="Lycopersicon esculentum"
 /db_xref="taxon:4081"
 gene 121..2991
 /gene="LHA1"
 CDS 121..2991
 /gene="LHA1"
 /BC_number="3.6.1.35"
 /codon_start=1
 /product="H+-ATPase"
 /protein_id="AAA34173.1"
 /db_xref="GI:170464"
 /translation="MAEKPEVLDVAVIKETVDLENPIEVEFENLRCTREGTLATAOE
 RLSTFGYNLEEKESKRLKLGFMNPLSWMEAAALMAALALANGGKPPMDQFVG
 IITLLINSTISFIEENNAGNAALMARLAPKAVLRDQKDEDAVLPDGIISI
 KGLDILPADARLEGGDPLKIDQSALTGESLPYTKGPDGVYSGSTCKOGEIEAVAT
 GVHFEFGKAHLVDSITNGVHFOKVLTAIGNFICSLAVGMILEIIVMYIQRKYP
 GIDNLVLLIGGPIAMPVLSVMAIGSHRLAOGGATIKRMTAIEEMAGDVYCSDK
 TGTLTINKLTVDKLTILEVFANGVADVYVLMARASRTENDADIAIVGMADKEA
 RAGRTIRHLPENPDOKRTALTYLDEGKMHVSKGAPQITLNLHNSDTERRHVTV
 IDKFAERGLRSLGVAOEVPEGRKESAGCPWQIALPLEPRHDSAEITRRALNIG
 VVWKMTTGDQALIGKETGRRLGTMTPSSALICQTKDESIYALPIDELEKADGFA
 GVPEHKEIYKRLQARKHICGTMGSDGVNDAPALKADIGIADVDATDARSASDVL
 TEGSLVITSAVTSRAIFQRMKNYITVAITIRIVLGFMALIMKRPDPPEVVL
 IALINDGTIMTISKDRVKRSPIDSKLAIEFTTGVGLAMMTVFFMAAYTTF
 FRIRGVSTLEKATIDDERKLASATILYOSTISQALIFITRSRWSFVERPDLVFA
 FVVAOLVATLIVAVYAWMSFALIEGIGMGVAVIWIYNTVYIPLDLIFLRVALSGK
 ANDVLEQRIATFRKDKFGLRELOMAHQRTLLGLQVDPKIFSETTFNELNQLA
 EEAKRRAELARLEHLTLKGHYESVYKGLDIEFTIQGSYTV"
 BASE COUNT 853 a 616 c 813 g 947 t
 ORIGIN Chromosome 3.
 Query Match 74.4%; Score 18.6; DB 15; Length 3229;
 Best Local Similarity 84.0%; Pred. No. 7.4e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 ggcgcagcagtgccatgcgcagcatt 25
 ++++++
 Db 502 GTGCAGCTGCTGCATTCGCAGCAT 478
 RESULT 35
 TOBPMAB/c 3233 bp DNA PLN 27-APR-1993
 LOCUS Nicotiana plumbaginifolia plasma-membrane H+ ATPase (pma3) gene,
 DEFINITION exons 1-9.
 ACCESSION M80491.1 GI:170296
 VERSION M80491
 KEYWORDS Plasma membrane ATPase.
 SOURCE Nicotiana plumbaginifolia (Library: EMBL4 genomic) DNA.
 ORGANISM Nicotiana plumbaginifolia
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 REFERENCE 1 (bases 1 to 3233)
 AUTHORS Perez,C., Michellet,B., Ferrant,V., Bogaerts,P. and Boutry,M.
 TITLE Differential expression within a three-gene subfamily encoding a
 plasma membrane H+-ATPase in Nicotiana plumbaginifolia

JOURNAL J. Biol. Chem. 267, 1204-1211 (1992)
 MEDLINE 92112762
 FEATURES Location/Qualifiers
 source 1..3233
 /organism="Nicotiana plumbaginifolia"
 /db_xref="taxon:4092"
 /tissue_1lb="EMBL4 genomic"
 join(578..895,988..1107;1636..1734,1826..1963,2102..2236,
 2325..2504,2605..2811,2906..3025,3122..3233)
 mRNA /partial
 /gene="pma3"
 /product="Plasma membrane H+ ATPase"
 exon 578..895
 /gene="pma3"
 /number=1
 CDS 782..799
 /note="upstream short ORF"
 /codon_start=1
 /protein_id="AAA34100.1"
 /db_xref="GI:170298"
 gene join(842..895,988..1107,1636..1734,1826..1963,2102..2236,
 2325..2504,2605..2811,2906..3025,3122..3233)
 CDS join(842..895,988..1107,1636..1734,1826..1963,2102..2236,
 2325..2504,2605..2811,2906..3025,3122..3233)
 /partial
 /gene="pma3"
 /codon_start=1
 /product="Plasma membrane H+ ATPase"
 /protein_id="AAA34099.1"
 /db_xref="GI:553114"
 intron /db_xref="GI:553114"
 /translation="MGEKPEVLDVAVIKETVDLENPIEVEFENLRCTREGTLATAOE
 RLSTFGYNLEEKESKRLKLGFMNPLSWMEAAALMAALALANGGKPPMDQFVG
 IITLLINSTISFIEENNAGNAALMARLAPKAVLRDQKDEDAVLPDGIISI
 KGLDILPADARLEGGDPLKIDQSALTGESLPYTKGPDGVYSGSTCKOGEIEAVAT
 GVHFEFGKAHLVDSITNGVHFOKVLTAIGNFICSLAVGMILEIIVMYIQRKYP
 GIDNLVLLIGGPIAMPVLSVMAIGSHRLAOGGATIKRMTAIEEMAGDVYCSDK
 TGTLTINKLTVDKLTILEVFANGVADVYVLMARASRTENDADIAIVGMADKEA
 RAGRTIRHLPENPDOKRTALTYLDEGKMHVSKGAPQITLNLHNSDTERRHVTV
 IDKFAERGLRSLGVAOEVPEGRKESAGCPWQIALPLEPRHDSAEITRRALNIG
 VVWKMTTGDQALIGKETGRRLGTMTPSSALICQTKDESIYALPIDELEKADGFA
 GVPEHKEIYKRLQARKHICGTMGSDGVNDAPALKADIGIADVDATDARSASDVL
 TEGSLVITSAVTSRAIFQRMKNYITVAITIRIVLGFMALIMKRPDPPEVVL
 IALINDGTIMTISKDRVKRSPIDSKLAIEFTTGVGLAMMTVFFMAAYTTF
 FRIRGVSTLEKATIDDERKLASATILYOSTISQALIFITRSRWSFVERPDLVFA
 FVVAOLVATLIVAVYAWMSFALIEGIGMGVAVIWIYNTVYIPLDLIFLRVALSGK
 ANDVLEQRIATFRKDKFGLRELOMAHQRTLLGLQVDPKIFSETTFNELNQLA
 EEAKRRAELARLEHLTLKGHYESVYKGLDIEFTIQGSYTV"
 exon 896..987
 /gene="pma3"
 /number=1
 intron 988..1107
 /gene="pma3"
 /number=2
 exon 1108..1635
 /gene="pma3"
 /number=2
 exon 1636..1734
 /gene="pma3"
 /number=3
 intron 1735..1825
 /gene="pma3"
 /number=3
 exon 1826..1963
 /gene="pma3"
 /number=4
 intron 1964..2101
 /gene="pma3"
 /number=5
 exon 2102..2236
 /gene="pma3"
 /number=5
 intron 2237..2324
 /gene="pma3"
 /number=5
 exon 2325..2504
 /gene="pma3"
 /number=6
 intron 2505..2604
 /gene="pma3"
 /number=6
 exon 2605..2811

ERACIREVHFLPVPVDRKATLTYIDAGNMHRSKAGPEOILTLCKRDKVKKAH
 GYIEKADRGRLSIAVAROEPEKESGPMOVLPLPDPDRHDSATIKRALN
 LGVNMKITGDOALAGKETGRKRLGNTMYSSLSLGGDKNVNAAGVPLDIKADG
 FAGVPERKIELVKRLQDRKHCGMTGSDVNDAPALKRADIGIADVDTRKASDI
 VTEPGLSVILISAVLTSRAIFQRMKNYITVAISITRVFGFMIALIMKDFEPMY
 LITALLNGITITISKDRVPSPLPDSWKLEIFATGIVIGYOAIMVVFPMVVD
 TFEVKEFVKPLTDGOMMAALYLOVASISOLAIETVRSRSPARPGMLMIGAVVA
 OLIALIVYANMSFKIIEGMMGMAIWMVITLTVYIPLDLKFTIRYALSGRAMN
 LLDNKTATTKDKYKEREEREAOMAAORTMGLORPETTNLPKESNRELSEIAEOA
 KRRAVARRELHLKGVESVYKLGIDITIOOHTV"

BASE COUNT 951 a 643 c 864 g 968 t
 ORIGIN

Query Match 74.4%; Score 18.6; DB 14; Length 3426;
 Best Local Similarity 84.0%; Pred. No. 7.3e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcacgcagcatt 25
 ||||| ||||| ||||| |||||
 Db 614 GCCAGCTGCTGCATTACCGACATT 590

RESULT 38
 AF179442/c
 LOCUS
 DEFINITION Lycopersicon esculentum plasma membrane H+-ATPase isoform LHA2 (LHA2) gene, complete cds; and 18S ribosomal RNA gene, partial sequence.
 ACCESSION AF179442
 VERSION AF179442.1 GI:5901756
 KEYWORDS
 SOURCE tomato.
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
 1 (bases 1 to 10302)
 Ewing,N.N., Wimmers,L.E., Meyer,D.J., Chetelat,R.T. and Bennett,A.B.
 TITLE Molecular cloning of tomato plasma membrane H+-ATPase
 JOURNAL Plant Physiol. 94, 1874-1881 (1990)
 REFERENCE 2 (bases 1 to 10302)
 AUTHORS Ewing,N.N. and Bennett,A.B.
 TITLE Assessment of the number and expression of P-type H(+)-ATPase genes in tomato
 JOURNAL Plant Physiol. 106 (2), 547-557 (1994)
 MEDLINE 95083753
 PUBMED 7991683
 REFERENCE 3 (bases 1 to 10302)
 AUTHORS Ro,S., Idate,R.R. and Ewing,N.N.
 TITLE The promoter of plasma membrane H+-ATPase LHA2 directs reporter gene expression in Arabidopsis in a pattern that suggests a role for this isoform in auxin-regulated processes
 JOURNAL unpublished
 AUTHORS 4 (bases 1 to 10302)
 Ro,S., Idate,R.R. and Ewing,N.N.
 TITLE Direct Submission
 JOURNAL Submitted (22-AUG-1999) Biological Sciences, California State University, Sacramento, 6000 J Street, Sacramento, CA 95819-6077, USA

FEATURES
 source
 1..10302 Location/Qualifiers
 /organism="Lycopersicon esculentum"
 /cultivar="VFNT Chery"
 /db_xref="taxon:4081"
 /chromosome="6"
 /map="between CSC83 and TG162"
 1..2146
 promoter
 1..2146
 /gene="LHA2"
 1..9460
 /gene="LHA2"
 join(2147..2218,2347..2421,2517..2636,3006..3104,

3191..3328,3475..3609,3700..3879,3976..4182,4302..4421,
 4505..4624,4705..4827,4918..5022,5229..5373,5992..6156,
 6269..6507,6694..6726,6817..6898,7480..7640,8113..8298,
 8379..8561,8646..8824,9084..9460)
 /gene="LHA2"
 /product="Plasma membrane H+-ATPase isoform LHA2"
 /evidence="experimental"
 join(2368..2421,2517..2636,3006..3104,3191..3328,
 3475..3609,3700..3879,3976..4182,4302..4421,4505..4624,
 4705..4827,4918..5022,5229..5373,5992..6156,6269..6507,
 6694..6726,6817..6898,7480..7640,8113..8298,8379..8561,
 8646..8824,9084..9180)
 /gene="LHA2"
 /codon_start=1
 /product="Plasma membrane H+-ATPase isoform LHA2"
 /protein_id="AA055399.1"
 /db_xref="GI:5901757"
 /translation="MGKPEVDAVLKETVDLENIPLEVEFENLRCTEGLTGTAOE
 RLALFGYVKLEEKESKFLKFLGFMNPLSLWVEAAIYALALANGKPPMDQVFG
 IITLVINSTISFLENNAGNAALALAPRAKYLROGKNEEDAAVLVDDISI
 KIGDILPADARILEGDKIDQSLNGESLPVTKPGDGYSGSTCKOGELEAVIAT
 GVHTFFGKAHLVDSTNOVGFQKVLTAIGNEFCISAVGMILEIITVPTQKRRP
 GIDMLLVLLIGLPIAMPVLSVTMAIGSHRLAQOGAITKRMTAIEEMAGMDVCSDK
 TGTLLINKLTVDKNLVEFAKGVADADIVVLAARASPTENODAIDRAIVGMLSDPEA
 RAGIREHTLPENFTDKRTALTYLDGKKHRYSKGAPEDILMLAHKSIDIERHAY
 IDKFAERGLSIVAYOEVPEGRKESGGMWFGILPDPDRHDSATIKRALNIG
 VNMKITGDOALAGKETGRKRLGNTMYSSLSLGGDKNVNAAGVPLDIKADG
 GVPEHKEIYKRLQARKHICGMDGVNDAPALKRADIGIANDATDARASDYL
 TEPGLSVISAVLTSRAIFQRMKNYITVAISITRVFGFMIALIMKDFEPMY
 IAIINDGITITISKDRVPSPLPDSWKLEIFATGIVIGYOAIMVVFPMVVD
 PPRVGVSTLQRTADDFRKLASAIYLOVSTISOLALFVRSRSPARPGMLMIGAVVA
 FLIOLAVTIAVYANMSFKIIEGMMGMAIWMVITLTVYIPLDLKFTIRYALSGR
 AMDLVLEORTAIFTRKDKPKGKEORLEOAHORLHGIQVDFTLFSATNFNLQDA
 EAKRRRAEIRARELHLKGVESVYKLGIDITIOOHTV"
 10212..>10302
 /product="18S ribosomal RNA"
 1596 c 1989 g 3676 t 1 others

BASE COUNT 3040 a 1596 c 1989 g 3676 t 1 others
 ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 10302;
 Best Local Similarity 84.0%; Pred. No. 6.3e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcacgcagcatt 25
 ||||| ||||| ||||| |||||
 Db 3299 GAGCAGCTGCTGCATTGCCAGCATT 3275

RESULT 39
 AC017995/c
 LOCUS
 DEFINITION Drosophila melanogaster. *** SEQUENCING IN PROGRESS ***. In ordered pieces.
 ACCESSION AC017995
 VERSION AC017995.1 GI:6553195
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 33204)
 Adams,M. and Venter,J.C.
 TITLE Direct Submission
 JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
 COMMENT
 This sequence was identified as CDM:10212937 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a "working draft" sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source Location/Qualifiers
 1..33204
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"

BASE COUNT 8981 a 7720 c 7703 g 8800 t

ORIGIN

Query Match 74.4%; Score 18.6; DB 65; Length 33204;
 Best Local Similarity 84.0%; Pred. No. 5,3e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcacgcacgcatc 25
 ||| ||||| ||||| ||||| |||||
 Db 3374 GCGTAGCAGCTGCACCGCCACCAT 3350

RESULT 40
 ATE27H5/C
 LOCUS ATE27H5 88548 bp DNA PLN 14-APR-2000
 DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone F27H5.
 ACCESSION AL163852
 VERSION AL163852.1 GI:7576197
 KEYWORDS
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 88548)
 Rieger, M., Mueller-Auer, S., Zipp, M., Schaefer, M., Mewes, H.W.,
 Rudd, S., Lemcke, K., Meyer, K.F.X., Queller, F. and Salanoubat, M.
 unpublished
 2 (bases 1 to 88548)
 EU Arabidopsis sequencing project.
 Direct Submission
 Submitted (14-APR-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
 lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de Project
 Coordinator: Marcel Salanoubat and Francis Queller, Groupement
 d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
 Gaston Cremieux, BP191, 91006 Evry Cedex, France;
 http://www.genoscope.cns.fr
 Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.

COMMENT

FEATURES
 source Location/Qualifiers
 1..88548
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="3"

misc_feature
 1..28386
 /note="This BAC overlaps with BAC T209. Please refer to
 EMBL accession AL138658 for analysis and annotation."
 complement(29790..30794)
 /gene="F27H5_10"
 29790..30794
 /gene="F27H5_10"
 complement(29790..30794)
 /gene="F27H5_10"
 /note="A probable sequencing error has led to the change
 in ORF which changes the first 5 amino acids (compared to
 AF132014). Genoscope has been asked to check."
 /codon_start=1
 /product="RING-H2 zinc finger protein ATL4"
 /protein_id="CAB87859.1"
 /db_xref="GI:7576196"
 /translation="MESLNPSSGGGNDSSSSDLSKPSVLIILIMLLILISVS
 ICFLLRNCNSHRVPLPSSSSVATVDSRRSGHRSVPETERSVYDIPKFK
 SSVTRSSSNSSGCAVLSKFEEDOLRLPLCCHAFADCDIDVLRNCPCLRS
 PLFASIEDLKLKSLAVGSSNGGENSEFRLEIGISRRROTPIPESEYDQHRYSIGFD
 YIVDDVSEISESNFRKQEDATTTTATATATVTTPTSEASLADIQNDGSRWLK

CDS

exon complement(29790..30794)
 /gene="F27H5_10"
 /number=1
 36842..37038
 /gene="F27H5_20"
 /number=1
 36842..37893
 /gene="F27H5_20"
 join(36842..37038,37144..37256,37378..37512,37638..37699,
 37777..37893)
 /gene="F27H5_20"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB87860.1"
 /db_xref="GI:7576199"
 /translation="MSYNOSRPDKSETQYRRTRGSTGQOQOQHSSSAGYKGAG
 APGSAAPASTYPDNSLSNRSFKPGNAOGGQPRVLPVNHNNNNPNASHRS
 OGEPIVGPTNPETSEFNNTGPIPKAPTSOSTVMSKINTEPNTAKSGDASQAFPLQ
 FSLGPDLMVPAFTTSAPPNMDQKRAQVEHLVLCICRYCDCCLHK"
 37039..37143
 /gene="F27H5_20"
 /number=1
 37144..37256
 /gene="F27H5_20"
 /number=2
 37257..37377
 /gene="F27H5_20"
 /number=2
 37378..37512
 /gene="F27H5_20"
 /number=3
 37513..37637
 /gene="F27H5_20"
 /number=3
 37638..37699
 /gene="F27H5_20"
 /number=4
 37700..37776
 /gene="F27H5_20"
 /number=4
 37777..37893
 /gene="F27H5_20"
 /number=5
 38085..44750
 /gene="F27H5_30"
 38085..39449
 /gene="F27H5_30"
 /number=1
 join(38085..39449,39541..42284,42684..42878,42965..43227,
 44009..44026,44321..44409,44508..44590,44687..44750)
 /gene="F27H5_30"
 /note="similarity to eukaryotic protein synthesis
 initiation factor, Homo sapiens, EMBL:AF104913"
 /codon_start=1
 /product="protein synthesis initiation factor-like"
 /protein_id="CAB87861.1"
 /db_xref="GI:7576200"
 /translation="MTRANIVLMOQSSLRTASNVPAVPRKSSNGGADNLMRKES
 HNPSEKADIVPHIAPPSQKSPITNIRMPVSOTPYOHTQVPHVPHFGGPNHMQT
 PVATISFQMPMPALMSGNTPOIPPOVYQGHPRPHNHGMIHQAGHGATPMQO
 IHPOLGHVGVGISPOYPOOGKYGAGAKTTPVKITPHDTHEELRLDRGRVSEGS
 TALKRHSNPPRRSQSVSFAPRPVNIQOVSNTMTIYRPVSYLNGGNSADAPRY
 HVPYIDSGRQLINQPHRTAPOLIRPAARHLSDSTSSYKARNADONMSSALPVNA
 KVSVPAGVSEKLSPPKDRSHGEVNIISQKNVACSLSSQDPSFVSQVPSASAP
 PAKSPVETVPLAKSSVETPPVKSVEETAPVTTEIRAEVNSISVEQDTKEP
 NHTLENROQTFMPLSDVSDPETATVAKENISLPTNCFRKQLKVTSTSDAPSDSDY
 TSIDKSTGSSHASSEISGSPQEKDKDKRTASDKLDEKSYISDAKHETLSGLVK
 AONEVDGATGVQSEKLAVDODSSDLPHSTHYLSVPLGHSFTKHSALVETTRN
 TSTGKKIKRIKILKADAGCTSPDMAVYQKPEKSSNVYHVSOMNLLPATPOAV
 EATVDPEVKNPEDEMDADVSTPKLETADNSNAKRGSSDEVDNCKINTERKYSKD
 FLKAFADCTALPEGFVSPDIALNALIVAYGASHHBDSPYTPGKVMQDQASGARLD

CDS

RRPSNVAGDDBRMTKNOGSLPAGYGGNNGFRGQGNNGSVLNPMPMOGPIISRPMPVGS
PMGSGRNTPLDERWGRSNGFOOKGLPSPPTPMQVMKARKKOVGTIAEDAKOR
OLKSLINKTLPONFEKILPEOVKSVNIDNAVTLGVISOIFPKALMEPEADFCF
HLSGALPDENNGEKITFKRLILNKKCOEFERKEEERARVAEEGQTEEMER
KRLQYRRMLGNIRLIGELLYKKRMTEKIMACLOKILGYQODPHENETKLMEST
IGVMIIDHNKAKFQMDGYFEKMMKMSCOELSSRVFPLINAIDLKKNWQSRMKEVG
KRIIEVHDHDAAROTQANRLSRGSMVSPORSGHMEFSPRGGGLSPDAOMGYS
HGPQGRGFSNODIFEDDRPSYEPMPMPORSCDEPITLGPQGLGQMSIRRPAY
ASNTYOSDATOAGGDSRRPAGGLNGFGSHSPASVYTRSSPOEGRGAYVHREPA
SRASDLSPEVSARQVLOGPSATVNSPRENLSEPOLNLSLAIKEYVSARDENETG
MCKMDNSPAHPTMISLWYDSEFRKREKEDLAKLVLNLYKSNDANLNFQVLKCF
ESVLKTLIEDAVNDAPKAAEFUGRLFGKSVTERVVTLEIGRLIDEGGEPSLLIEPL
GGDYVSLVEMIKTEAGEETLVEIRSSGEENEDEGTGASIRQIKKMEHOSKVF
CEFCRGYGVKRAVIGICGKDCGKAGCAVTMTASAVTVSTIRLRLEQIEG"
39450. .39540
/gene="F27H5_30"
/number=1
39541. .42284
/gene="F27H5_30"
/number=2
42285. .42683
/gene="F27H5_30"
/number=2
42684. .42878
/gene="F27H5_30"
/number=3
42879. .42964
/gene="F27H5_30"
/number=3
42965. .43227
/gene="F27H5_30"
/number=4
43228. .44008
/gene="F27H5_30"
/number=4
44009. .44026
/gene="F27H5_30"
/number=5
44027. .44320
/gene="F27H5_30"
/number=5
44321. .44409
/gene="F27H5_30"
/number=6
44410. .44507
/gene="F27H5_30"
/number=6
44508. .44590
/gene="F27H5_30"
/number=7
44591. .44686
/gene="F27H5_30"
/number=7
44687. .44750
/gene="F27H5_30"
/number=8
44751. .46962
/gene="F27H5_40"
/complement(join(45714. .45844,45993. .46158,46255. .46364,
46453. .46671,46758. .46962))
/gene="F27H5_40"
/complement(45714. .45844)
/gene="F27H5_40"
/number=1
complement(join(45714. .45844,45993. .46158,46255. .46364,
46453. .46671,46758. .46962))
/gene="F27H5_40"
/codon_start=1
/product="regulatory subunit of protein kinase CK2"
/protein_id="CAB87862.1"
/db_xref="GI:7576201"
/translation="MYKERSGGGGSSSEILIGALIDRKRIINALNKKLEKSSSTSTT
TSRVSSKDKDPFSFTSTKQLPDVESETDSEGDVSGEGDPTSMISMFNLRGNDP

Query Match 74.4%; Score 18.6; DB 13; Length 88548;
Best Local Similarity 84.0%; Pred. No. 4,7e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy 1 ggcgcagcagtgcatcgccagcatt 25
Db 74472 GAGCAGCGTGCATTTCACGATT 74448
RESULT 41
ATT8B10/C
LOCUS 103787 bp DNA PLN 20-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone T8B10.
ACCESSION AL138646
VERSION AL138646.2 GI:7287982
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 103787)
Rieger,M., Mueller-Auer,S., Zipp,M., Schaefer,M., Mewes,H.W.,
Lemcke,K., Mayer,K.F.X., Quetier,F. and Salanoubat,M.
Unpublished
2 (bases 1 to 103787)
EU Arabidopsis sequencing project.
Direct Submission
Submitted (20-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de project
Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
d'Interet Public, Centre National de Sequenage - GENOSCOPE; 2 rue
Gaston Cremieux, BP191, 91006 Evry Cedex, France;
http://www.genoscope.cns.fr
On Mar 22, 2000 this sequence version replaced gi:6899957.
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
Location/Qualifiers
1. 103787
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
complement(8353. .8556)
/number=1
8353. .12143
/gene="T8B10.10"
complement(join(8353. .8556,8641. .8947,9034. .9209,
9288. .9351,9485. .9579,9673. .9688,10013. .10144,
10459. .10539,10632. .12143))
/note="similarity to arm repeat containing proteina
homolog, Arabidopsis thaliana, EMBL:AL133314"
/codon_start=1
/product="Arm repeat containing protein-like"
/protein_id="CAB81821.1"
/db_xref="GI:7287983"
/translation="MSRRVQRQREDNGKRVKSPSYTVIGVEDLAPKVOQYVMSTSLP
YDYFHLFTRLNRYDRASLASCTRTWSLASFMSGLDPLRAHEDLSMAASLARC
VDLOKIRRGVSDADAIHLKASLSEISDYCRKITDARLSMARHLELSIOGP
DPCRRLTSDAIRVAFRCPCPKLRLRVSGMDVSSPAESLAKCPOLSDGPGDCINT
NEALAGVYSIRYLSVAGTSNIRKKALEWMEKPLRLIGDVSRTTIDHIAVRLKS
SOLUKVLCALNCPYLEDKSYSSNRFKGVLAFTDPEELASTIFADNSKKNIFS
YMRLLKDKSDIDEIMLMIMWISHTLRAESSNSQGLDPLNGOATLLSLMOSA
QEDVQERAAATGLATFIYVDDENASIDCGRAEAVMPRGAILLELAKSMEEGLOESA
KAINALSVNAKVAKVAVEGGISVLDPLASNMRLVAEEAAGGLMINSVEEHRNIA
OAGGVNAYLDLIFPMPCGCVLEERAAGALANLAADKCSMEVARRAGVAYLVMLRN
CKYRGQEOQAARALANLAAGDSNGNNAVGOAGALEALVOLQTQSPHGVKEAAGA
LWNLAFDQKNRSTAAFGVAVLALAKSSNSSTGLOEVAGALGLSVSENSTAI
GHEGCIPLIALVSEADVETAAAGLWMLSPNPGVALRIVEGGVVALVOLCSSSV
SKMARFMAALALAYMFGDMDEYAMIGTSLSTSKSVTLNGARTMALDOIKAIFKTFM

EHOJFSTGLALSASAPLMSIAQSEAPRIPEGALHCSSTSEGREVTMLRNPCVLIRACAA
 FALLOFTIPREBRHMHIAISLQNAQNGEARGLRSAAAASMPRAKILFMIVLRIENHOC
 AESECGKKVSYNR."
 complement(8537. .8640)
 /number=1
 complement(8641. .8947)
 /number=2
 complement(8948. .9033)
 /number=2
 complement(9034. .9209)
 /number=3
 complement(9210. .9287)
 /number=3
 complement(9288. .9351)
 /number=4
 complement(9352. .9484)
 /number=4
 complement(9485. .9579)
 /number=5
 complement(9580. .9672)
 /number=5
 complement(9673. .9888)
 /number=6
 complement(9889. .10012)
 /number=6
 complement(10013. .10144)
 /number=7
 complement(10145. .10458)
 /number=7
 complement(10459. .10539)
 /number=8
 complement(10540. .10611)
 /number=8
 complement(10632. .12143)
 /number=9
 14024. .15549
 /gene="T8B10.20"
 complement(join(14024. .14116,14198. .14297,14361. .14461,
 14632. .14731,14825. .14918,15040. .15131,15138. .15408,
 15494. .15549))
 /note="similarity to CGI-94 protein, Homo sapiens,
 EMBL:AF151852"
 /codon.start=1
 /product="putative protein"
 /protein_id="CA881822.1"
 /db_xref="GI:7287984"
 /translation="MSSLRALVIRPAHKERSQPEARKRFGILEKHNDVILIRANAYHKKK
 QETLKIILROAAKNDPEFTKINSKTYDGRHPRKDEVKYSABELIMKTDQIGVV
 FOKMSKKNKIDLTFTSLQGTGSSRRHYVVAEDREARELEVGGRSKSPISVEILP
 KDIKKMDRSYRLERKSKSAKLELYTDMQMEIDKKGKKRLRDELINNGKAA
 VYKRRARRKK"
 complement(14024. .14116)
 /number=1
 complement(14117. .14197)
 /number=1
 complement(14198. .14297)
 /number=2
 complement(14298. .14380)
 /number=2
 complement(14381. .14461)
 /number=3
 complement(14462. .14651)
 /number=3
 complement(14652. .14731)
 /number=4
 complement(14732. .14824)
 /number=4
 complement(14825. .14918)
 /number=5
 complement(14919. .15039)
 /number=5
 complement(15040. .15131)
 /number=6

intron	complement(15132. .15317)
exon	/number=6 complement(15318. .15408)
intron	/number=7 complement(15409. .15493)
exon	/number=7 complement(15494. .15549)
gene	/number=8 16547. .17851
CDS	/gene="rB10_30" complement(join(16547. .16646,17116. .17228,17358. .17467 17581. .17851)) /product="hypothetical protein" /codon_start=1 /protein_id="CAB81823.1" /db_xref="GI:7287985" /translation="MKRFAILDQNFSPFPPDSCKAMSPNCSLSRRSLVLYLVASPC LLPLPAISSSKTKRKSPYDERLLQNKRIQRENNAPDERPFRGFEVKKYASDNNQ IKDSGILIRDNVQGGDGPKKDGGQVFFHYIGVNESGRIRIDSTYIQGSPARIRMTNTN LVPGNKQFEVDEVLSTIONCERTTIGFYSVTCS" complement(16547. .16646) /number=1 complement(16647. .17115) /number=1 complement(17116. .17228) /number=2 complement(17229. .17357) /number=2 complement(17358. .17467) /number=3 complement(17468. .17580) /number=3 complement(17581. .17851) /number=4 18460. .20691 /gene="rB10_40" complement(18460. .20691) /note="Contains ATP-dependent DNA ligase signatures

```

/codon_start=1
/product="putative protein"
/protein_id="CAB81824.1"
/db_xref="GI:7287986"
/translation="MASNPYTKRRSPNVVPPDPYKSGGGGFCKSVLFALEFL
ALIPSPQAPFVEETLYTFKMLHLLFGLVAAGLESRNVAEADLMTREDEE
SLSYRYLFOVSVDFEEDNOCSEPHLDGSDVAARSVAKSEKSEPELLE
SFEQGTNRVAMNSQYCKGKSKVYVAPAGLDGHVHQLPLPIRLRSSLRMAAL
DKSRADSCDAVNAAEASLLADNFEVYLAASVYPQAPPEMAGTGNPSRQD
ISVDLTLISRRSTGSSSOTSTASIONKRESPSSVAESLNSNVELKERSRO
ESSRSPSLPPSLPLSPPELVPNDTRRSLVTDPTPRASHSRHSDSLTSL
EVRRGFENLEKSGVGRAAEFKPKERKSLRLAAESSRGKSRHRYVPSLIS
PVGADDSSTVTRDLOOLKRNHLEENIRKGVADHNNLRVKGGRSHDLETTED
KDEKSEFPALPDPYPTNANKSRAMSESGRGPTPEKRYVTRSLDTPYSDSK
RKKLDPGDKEMKLDGRIEPRSRRASSNVSLGKGVYTRISDRNGKQVKTGDSSE
RAEAAVWESGSRRRROEELSLVLAHQESSSTRKSPPEVNAEEQAPQDPEVTE
EEEEEAAMESSNASHNDEKRAAGETAKFREQIRLOKLISGEOPGGGTGIFRNS
QFR"
/complement(18460..20691)
/number=1
/2335..23917
/gene="H8B10_50"
/complement(join(22335..22644,23038..23117,23202..23560,
23719..23917))
/note="Contains 'Homeobox' domain signature and profile
AA192-215"
/codon_start=1

```

```
Query Match      74.4%; Score 18.6; DB 14; Length 103787;
Best Local Similarity 84.0%; Pred. NO. 4.Ge+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

Db 1019 GAGCAGCAGCTGCATTTCACGATT 995

RESULT 42
ATT12K4 108879 bp DNA PLN 05-APR-2000
LOCUS Arabidopsis thaliana DNA chromosome 3, BAC clone T12K4.
ACCESSION AL138640
VERSION AL138640.1 GI:6899910
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 108879)
Montfort, A., Casacuberta, E., Puigdomenech, P., Mewes, H. W., Rudd, S., Lemcke, K., Mayer, K. F. X., Quettier, F. and Salanoubat, M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 108879)
EU Arabidopsis sequencing project.
AUTHORS Direct Submission
TITLE Submitted (05-APR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemekemips.biochem.mpg.de, mayeremips.biochem.mpg.de, Project Coordinator: Marcel Salanoubat and Francis Quettier, Groupement d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France; http://www.genoscope.cns.fr
COMMENT Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES
source location/Qualifiers
1..108879
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
1..5515
/note="overlap to BAC T32A11, please refer also to EMBL:AL138653 for analysis and annotation"
1099..5822
/note="copla-like LTR retrotransposon"
1099..1395
/note="297 bp long terminal repeat"
5526..5822
/note="297 bp long terminal repeat"
6818..7357
/note="28 bp long motif tgatttcaggtttgagcgaataata with up to 9 mismatches repeated"
9190..9239
/note="50 bp at tandem repeat"
10223..10372
/gene="T12K4_10"
/number=1
10223..10718
/gene="T12K4_10"
join(10223..10372,10433..10504,10644..10718)
/gene="T12K4_10"
/note="similarity to crpl protein - Zea mays, PIR:T01685"
/codon_start=1
/product="putative protein"
/protein_id="CAB86439.1"
/db_xref="GI:7523381"
/translation="MENEGIEPNLIMSNVLVNAFGTAGSTWKLSEFIILKRLVYLLRAIKLIGYTPDVVYTYSLMKAFTRPKKYEKGSGSWVYGRSEKTAIAKCFYGS"
10373..10432
/gene="T12K4_10"
/number=1
10433..10504
/gene="T12K4_10"
/number=2
10505..10643

exon /gene="T12K4_10"
/number=2
10644..10718
/gene="T12K4_10"
/number=3
19016..20305
/gene="T12K4_20"
join(19016..19120,19958..20305)
/gene="T12K4_20"
/note="similarity to peroxidase - Lycopersicon esculentum, PIR:S32768"
/codon_start=1
/product="putative protein"
/protein_id="CAB86440.1"
/db_xref="GI:7523382"
/translation="MERKKESKEGILAAKEKRLQTNLVRLDRFIDSPSGDQPTLDSVEINIKRRLQISIKTSLEHCEPKOVSCSDVITLSARDVALTGPPLSVLGLRKDSLSTPSKHWADSEPPSTADVDITLSLFSNGMTIEDQSVAIMGYTL"
19016..19120
/gene="T12K4_20"
/number=1
19121..19957
/gene="T12K4_20"
/number=1
19958..20305
/gene="T12K4_20"
/number=2
20779..20814
/gene="T12K4_30"
/number=1
20779..25885
/gene="T12K4_30"
join(20779..20814,21087..21444,21675..21970,22038..22173,22321..22451,22560..22679,22767..22793,22894..22898,23085..23195,23291..23408,23487..23572,23659..23820,23917..23994,24076..24148,24412..24470,24617..24928,25268..25468,25583..25885)
/gene="T12K4_30"
/note="similarity to hypothetical proteins - Arabidopsis thaliana"
/codon_start=1
/product="putative protein"
/protein_id="CAB86441.1"
/db_xref="GI:7523383"
/translation="MYHRRRRLMVRPLFALDHYPTTKAKINAYSKAEYIGQIFALGTPKMTFLBESPGDLFRTPANKASFSGKLVYLGLICQLVKKCTEKMMYFPGPIPLGLREWARLTGLECGTYPRNSAIEAVMOQKEGTSFLTLDDCASFKEGEDIEKLKNGCKOOSYALHGPPGLGLPFAFETIPSIASLVSEKEDTVFERSIKNLAKLITKSTILACDAEKACTVAVVDVYIIVPVDSDCPKSLVMKREKODPVNDIILMLNATSGMRKILKKRRVDPIDIKRRKITNSDGEHSGSGDSRDEPFKYADONEGTRASOKSDSPSEKVENKRVKGGKKNKVOVSGOAEFTISLPIEELSSDDNEDPSNGSRSDDEM KDIAPTOPVRPNPFGNDGDEVSARKKHHTSYSDVMERTIYLGASPPHOEGSGSEIREDSKIHONVSDSVWEKEFDSVSPPLVTCIPORDINEBGLAOSKPLMSDVTSTHQDVAKDVGNTEESVVAELPKMBENSPEPHSQNEPRVDEIDVATHEDNKRNDPPTSGDGEVLKKAQSPSEFVGEVEDTADQPIVVPDRLLEAFDANLNINVAEKESEPIPEKDVPEHPGKKDDVVMGVSGTEENVAHVEKEKKVRKVFADDLKDGDISHTKSDVFERMERESTLNSDYLKAEADVPYNLEKVEDEKPEKILCSIMKEGMYLHNGENMKRNRIVAVDPLPWLRSRATARFSCVNPETFKDKFLDFPARGESINRPKLRMINDDSVRSQILPILEMLPYLVATCKDYLEPVLTPPTVYLRNQLSONPTTGDCSLYANNPIELMYLQNPMDLILIEBANMYNRYKGYAVDLNHHGIGIT"
20815..21086
/gene="T12K4_30"
/number=1
21087..21444
/gene="T12K4_30"
/number=2
21445..21674
/gene="T12K4_30"
/number=2
21675..21970
/gene="T12K4_30"
/number=3
21971..22037

intron

exon	/gene="T12K4_30"	/number=3	22038. .22173	/gene="T12K4_30"	22174. .22320	/gene="T12K4_30"	22321. .22451	/gene="T12K4_30"	22452. .22559	/number=5	22560. .22679	/gene="T12K4_30"	22680. .22766	/number=6	22767. .22793	/gene="T12K4_30"	22794. .22893	/number=7	22894. .22998	/gene="T12K4_30"	22999. .23084	/number=8	23085. .23195	/number=9	23196. .23290	/number=9	23291. .23408	/number=10	23409. .23486	/number=10	23487. .23572	/number=11	23573. .23658	/number=11	23659. .23820	/number=12	23821. .23916	/number=12	23917. .23994	/number=13	23994. .24079	/number=13	24079. .24164	/number=14	24164. .24249	/number=14	24249. .24334	/number=15	24334. .24419	/number=15	24419. .24504	/number=16	24504. .24589	/number=16	24589. .24674	/number=17	24674. .24759	/number=17	24759. .24844	/number=18	24844. .24929	/number=18	24929. .25014	/number=19	25014. .25099	/number=19	25099. .25184	/number=20	25184. .25269	/number=20	25269. .25354	/number=21	25354. .25439	/number=21	25439. .25524	/number=22	25524. .25609	/number=22	25609. .25694	/number=23	25694. .25779	/number=23	25779. .25864	/number=24	25864. .25949	/number=24	25949. .26034	/number=25	26034. .26119	/number=25	26119. .26204	/number=26	26204. .26289	/number=26	26289. .26374	/number=27	26374. .26459	/number=27	26459. .26544	/number=28	26544. .26629	/number=28	26629. .26714	/number=29	26714. .26799	/number=29	26799. .26884	/number=30	26884. .26969	/number=30	26969. .27054	/number=31	27054. .27139	/number=31	27139. .27224	/number=32	27224. .27309	/number=32	27309. .27394	/number=33	27394. .27479	/number=33	27479. .27564	/number=34	27564. .27649	/number=34	27649. .27734	/number=35	27734. .27819	/number=35	27819. .27904	/number=36	27904. .27989	/number=36	27989. .28074	/number=37	28074. .28159	/number=37	28159. .28244	/number=38	28244. .28329	/number=38	28329. .28414	/number=39	28414. .28499	/number=39	28499. .28584	/number=40	28584. .28669	/number=40	28669. .28754	/number=41	28754. .28839	/number=41	28839. .28924	/number=42	28924. .29009	/number=42	29009. .29094	/number=43	29094. .29179	/number=43	29179. .29264	/number=44	29264. .29349	/number=44	29349. .29434	/number=45	29434. .29519	/number=45	29519. .29604	/number=46	29604. .29689	/number=46	29689. .29774	/number=47	29774. .29859	/number=47	29859. .29944	/number=48	29944. .30029	/number=48	30029. .30114	/number=49	30114. .30199	/number=49	30199. .30284	/number=50	30284. .30369	/number=50	30369. .30454	/number=51	30454. .30539	/number=51	30539. .30624	/number=52	30624. .30709	/number=52	30709. .30794	/number=53	30794. .30879	/number=53	30879. .30964	/number=54	30964. .31049	/number=54	31049. .31134	/number=55	31134. .31219	/number=55	31219. .31304	/number=56	31304. .31389	/number=56	31389. .31474	/number=57	31474. .31559	/number=57	31559. .31644	/number=58	31644. .31729	/number=58	31729. .31814	/number=59	31814. .31899	/number=59	31899. .31984	/number=60	31984. .32069	/number=60	32069. .32154	/number=61	32154. .32239	/number=61	32239. .32324	/number=62	32324. .32409	/number=62	32409. .32494	/number=63	32494. .32579	/number=63	32579. .32664	/number=64	32664. .32749	/number=64	32749. .32834	/number=65	32834. .32919	/number=65	32919. .33004	/number=66	33004. .33089	/number=66	33089. .33174	/number=67	33174. .33259	/number=67	33259. .33344	/number=68	33344. .33429	/number=68	33429. .33514	/number=69	33514. .33599	/number=69	33599. .33684	/number=70	33684. .33769	/number=70	33769. .33854	/number=71	33854. .33939	/number=71	33939. .34024	/number=72	34024. .34109	/number=72	34109. .34194	/number=73	34194. .34279	/number=73	34279. .34364	/number=74	34364. .34449	/number=74	34449. .34534	/number=75	34534. .34619	/number=75	34619. .34704	/number=76	34704. .34789	/number=76	34789. .34874	/number=77	34874. .34959	/number=77	34959. .35044	/number=78	35044. .35129	/number=78	35129. .35214	/number=79	35214. .35299	/number=79	35299. .35384	/number=80	35384. .35469	/number=80	35469. .35554	/number=81	
------	------------------	-----------	---------------	------------------	---------------	------------------	---------------	------------------	---------------	-----------	---------------	------------------	---------------	-----------	---------------	------------------	---------------	-----------	---------------	------------------	---------------	-----------	---------------	-----------	---------------	-----------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	---------------	------------	--

pieces.
 AC014784
 AC014784.1 GI:6436551
 HG: HGSC_PHASE2.
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 120181)
 Adams, M. and Venter, J.C.
 Direct Submission
 Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 This sequence was identified as CDM:10209834 by the submitter.
 For further information on this sequence e-mail to fly@celera.com
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 Location/Qualifiers
 1..120181
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 33873 a 26853 c 26669 g 32786 t
 BASE COUNT
 ORIGIN

Query March	74.4%;	Score 18.6;	DB 63;	Length 120181;
Best Local Similarity	84.0%;	Pred. No. 4.5e+02;		
Matches 21;	Conservative	0;	Mismatches 4;	Indels 0;
				Gaps 0;
Oy	1	gcgcagcaggtgcatgcgcagcatt	25	
D3	32083	gcgcagcagctgcattccgcagcagct	32059	

RESULT	44
AC008221/c	
LOCUS	AC008221 184278 bp DNA INV 22-FEB-2001
DEFINITION	Drosophila melanogaster, chromosome 3R, region 100B-100C, BAC clone.
ACCESSION	BACR30B13 AC008221 complete sequence.
VERSION	AC008221.4 GI:13096037
KEYWORDS	HMG.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster

REFERENCE
AUTHORS

1 (bases 1 to 184278)
Celniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Bazon,J., Beeson,K.Y., Bisam,D.A.,
Carlson,L.W., Center,A., Chape,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorett,V., E. Galle,R.F., Gary,N.S., George,R.A.,
Rerterla,S., Fritze,E., Galle,R.F., Gary,N.S., George,R.A.,
Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Idegaw,C., Jajali,M., Kruse,D., Li,P., Matzel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
Paclet,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phoumenavong,S., Pittman,G.S., Puti,V., Richards,S., Scheeler,F.,
Stapleton,M., Strong,R., Svrtskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.

TITLE
JOURNAL
REFERENCE
AUTHORS

Sequencing of Drosophila chromosome 3R, region 100B-100C
Unpublished
2 (bases 1 to 184278)
Celniker,S.E., Agbayan,A., Arcaina,T.T., Baxter,E., Blazek,R.G.,
Batenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Kara,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomutan,A.B., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Paclet,J.M., Park,S.,

Search completed: October 9, 2001, 12:11:55
Job time: 3830 sec

2 (bases 1 to 186549)
Celniker, S. E., Abbaya, A., Arcina, T. T., Baxter, E., Blazek, R. G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Holt, R. A., Evans, C. A., Gockayne, J. D., Amanatides, P. G.,
Rogers, Y. A., Baldwin, D., Banazon, J., Beeson, K. Y.,
Carlson, J. W., Center, A., Champe, M., Davenport, L. B., Dietz, S. M.,
Dodson, K., Dorset, V., Dou, L. E., Doyle, C., Dresnek, D., Farfan, D.,
Ferreira, S., Frisbe, E., Galle, R. F., Garg, N. S., George, R. A.,
Gonzalez, M., Houck, J., Hoskins, R. A., Hostin, D., Howard, T. J.,
Ibbagyan, C., Jallali, M., Kruse, D. L. P., Mattei, B., Mosher, A.,
McIntosh, T. C., Moy, M., Murphy, B., Nelson, C., Nelson, K. A., Nunn, J.,
Paclet, J., Parags, V., Park, S., Patel, S., Pfeiffer, B.,
Phoumenavong, S., Pltman, G. S., Pui, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Svirskas, R., Tector, C., Williams, S. M.,
Zaveril, J. S., Smith, H. O., Rubin, G. M., and Venter, J. C.

Sequencing of *Drosophila* chromosome 3R, region 100B-100C

Unpublished

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:37:32 ; Search time 470.56 Seconds
(without alignments)
33.359 Million cell updates/sec

Title: US-09-396-196f-5

Sequence: 1 gcgcagcagtcgcatcgccagcatt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

N_Geneseq_0601:*

1: /SIDSI/gcgdata/geneseq/geneseqn/NA1980.DAT:*

2: /SIDSI/gcgdata/geneseq/geneseqn/NA1981.DAT:*

3: /SIDSI/gcgdata/geneseq/geneseqn/NA1982.DAT:*

4: /SIDSI/gcgdata/geneseq/geneseqn/NA1983.DAT:*

5: /SIDSI/gcgdata/geneseq/geneseqn/NA1984.DAT:*

6: /SIDSI/gcgdata/geneseq/geneseqn/NA1985.DAT:*

7: /SIDSI/gcgdata/geneseq/geneseqn/NA1986.DAT:*

8: /SIDSI/gcgdata/geneseq/geneseqn/NA1987.DAT:*

9: /SIDSI/gcgdata/geneseq/geneseqn/NA1988.DAT:*

10: /SIDSI/gcgdata/geneseq/geneseqn/NA1989.DAT:*

11: /SIDSI/gcgdata/geneseq/geneseqn/NA1990.DAT:*

12: /SIDSI/gcgdata/geneseq/geneseqn/NA1991.DAT:*

13: /SIDSI/gcgdata/geneseq/geneseqn/NA1992.DAT:*

14: /SIDSI/gcgdata/geneseq/geneseqn/NA1993.DAT:*

15: /SIDSI/gcgdata/geneseq/geneseqn/NA1994.DAT:*

16: /SIDSI/gcgdata/geneseq/geneseqn/NA1995.DAT:*

17: /SIDSI/gcgdata/geneseq/geneseqn/NA1996.DAT:*

18: /SIDSI/gcgdata/geneseq/geneseqn/NA1998.DAT:*

19: /SIDSI/gcgdata/geneseq/geneseqn/NA1999.DAT:*

20: /SIDSI/gcgdata/geneseq/geneseqn/NA2000.DAT:*

21: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:*

22: /SIDSI/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	20	AAAX01303
2	25	100.0	1084	10	AAAX01329
3	25	100.0	1121	7	AAAN60496
4	25	100.0	5872	15	AAQ62386
5	18.4	73.6	860	21	AAAF15143
6	18.2	72.8	2533	22	AAZ11737
7	18.2	72.8	2536	22	AAAF56388
8	17.8	71.2	300	21	AAA01491
9	17.8	71.2	1197	21	AAA64352
10	17.8	71.2	1197	21	AAA39343
11	17.8	71.2	2048	22	AAAF26295

C	12	17.6	70.4	40	17	AAAT31064
C	13	17.6	70.4	384	21	AAAC03291
C	14	17.6	70.4	1215	21	AAAB60774
C	15	17.6	70.4	1344	21	AAZ56381
C	16	17.6	70.4	2126	19	AAVA1999
C	17	17.6	70.4	2126	20	AAZ11174
C	18	17.6	70.4	2214	22	AAAF61051
C	19	17.6	68.8	423	18	AAAT80740
C	20	17.2	68.8	568	21	AAAF09857
C	21	17.2	68.8	769	21	AAAC58894
C	22	17.2	68.8	1462	13	AAQ31876
C	23	17.2	68.8	1462	14	AAQ53208
C	24	17.2	68.8	1815	18	AAAF6495
C	25	17.2	68.8	2850	20	AAZ24632
C	26	17.2	68.8	2856	21	AAAC65871
C	27	17.2	68.8	5532	21	AAA81538
C	28	17.2	68.8	349980	21	AAAF21611
C	29	17.2	68.8	837096	21	AAAB1489
C	30	17	68.0	468	21	AAA55225
C	31	17	68.0	514	21	AAAC49411
C	32	17	68.0	828	20	AAAX08449
C	33	17	68.0	828	20	AAAX27714
C	34	17	68.0	911	21	AAZ44719
C	35	17	68.0	952	20	AAAX08443
C	36	17	68.0	952	20	AAAX27708
C	37	17	68.0	2320	21	AAAF21879
C	38	17	68.0	2905	21	AAAC44710
C	39	17	68.0	32998	21	AAA55186
C	40	16.8	67.2	20394	22	AAZ22301
C	41	16.8	67.2	20394	22	AAAF24892
C	42	16.8	67.2	534720	19	AAV30458
C	43	16.8	67.2	536165	19	AAV30458
C	44	16.6	66.4	68	20	AAZ20565
C	45	16.6	66.4	70	20	AAZ20584

ALIGNMENTS

RESULT	1
AAAX01303	standard; DNA: 1041 BP.
ID	AAAX01303: standard; DNA: 1041 BP.
XX	AAAX01303:
AC	AAAX01303: (first entry)
XX	
DT	12-APR-1999
DE	E. coli biotin synthetase (biob) coding sequence.
XX	
KW	DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW	biotin synthetase; biotin production; vitamin H; Biob; ss.
XX	
OS	Escherichia coli.
XX	
PN	US5869719-A.
XX	
PD	09-FEB-1999.
XX	
PF	30-APR-1997; 97US-0846338.
XX	
PR	30-APR-1997; 97US-0846338.
PR	08-MAR-1995; 95US-0401068.
XX	
PA	(NOVS) NOVARTIS FINANCE CORP.
XX	
PI	Patton DA:
XX	
DR	WPI: 1999-152902/13.
XX	
PT	P-PSDB; AAW73906.
PT	Transgenic plants with high biotin levels - transformed with DNA encoding di-amino-pelargonic acid amino-transferase or biotin synthase

NE-KappaB binding
Human secreted pro
Human SENR ligand
Escherichia coli f
Human GPR14 (G-pro
Human GPR14 coding
P. putida KT2440-a
Type II topoisomer
Fusarium venenatum
Human tumour suppr
Cyclin D2 pseudoge
Human cyclin D2 ps
Human glucose tran
Human lung tumor a
Human lung cancer-
N. meningitidis pa
Neisseria meningit
N. meningitidis pa
C. symbiosum open
Arabidopsis thalia
Membrane bound por
Active membrane-bo
Bovine c-Kit GST-c
Porcine stem cell
Porcine stem cell
Human breast and o
Arabidopsis thalia
Centarchaeum symbo
cDNA encoding a hu
Pimaricin biosynth
Rhizobium species
Rhizobium species
CTLA-4 variable 11
CTLA-4 variable 11

XX Example 2: Column 37-40; 34pp; English.

XX This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthetase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.

SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgcatgcgcagcatt 25
 ||||||||||||||||||
 Db 79 gcgcagcaggtgcatgcgcagcatt 103

RESULT 2

AAAN91329 standard; DNA; 1084 BP.

AC AAAN91329;

DT 15-FEB-1990 (first entry)

DE E.coli Bio B gene.

DE E.coli Bio B gene; biotin.

OS Escherichia coli.

Key Location/Qualifiers

FT CDS 24..1064

PN GB216530-A.

PD 11-OCT-1989.

PF 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

PR 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISH.

PI Pearson BM, McKee RA;

DR WPI; 1989-295085/41. P-PDB P91392

PT plasmid contg. gene(s) for expression of biotin synthetase enzymes
 derived from E.coli and capable of replication and expression in other
 microorganisms, esp. yeast.

PS Table 3; page 33-4; 52pp; English.

CC The gene can be used in a plasmid for expression of enzymes of the biotin
 CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, PKV49 and pCK495, and plasmid pCK965 for
 CC lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.

SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgcatgcgcagcatt 25
 ||||||||||||||||||
 Db 102 gcgcagcaggtgcatgcgcagcatt 126

RESULT 3

AAAN60496 standard; DNA; 1121 BP.

AC AAAN60496;

DT 17-OCT-1991 (first entry)

DE Sequence encoding biotin synthesising enzyme.

DE Biotin synthetic enzyme; E.coli; deshydrobiotin; ds.

Key Location/Qualifiers

FT CDS 42..1082

PN JPE1149091-A.

PD 07-JUL-1986.

PF 24-DEC-1984; 84JP-0272605.

PR 24-DEC-1984; 84JP-0272605.

PA (NIPS) NIPPON SODA KK.

DR WPI; 1986-216622/33.

DR P-PDB; AAP60536.

PT Double stranded DNA encoding biotin synthesising enzyme -

PT comprises transformed mutant E.coli strain contg. cyclic doubled

PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

PS Disclosure; Page 534; 23pp; Japanese.

CC The sequence may be expressed by a transformed E.coli host, cultured

CC in a medium containing deshydrobiotin.

CC Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgcatgcgcagcatt 25
 ||||||||||||||||||
 Db 120 gcgcagcaggtgcatgcgcagcatt 144

RESULT 4

AAQ62386 standard; DNA; 5872 BP.

AC AAQ62386;

DT 16-NOV-1994 (first entry)

DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

KW Biotin; expression; enterobacteria; vitamin H; synthesis;
 plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;

KW promoter; pBac; biotin synthetase; KAPA synthetase;
 8-amino-7-oxononanoate synthetase; pimeloyl-CoA; DFB synthetase;

```

KW delthiobiotin synthase; DAPA synthase;
KW S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
XX seborrhoea; dermatitis; ds.
XX Escherichia coli DSM498.
FH Key Location/Qualifiers
FT promoter 1..96
FT /tag= a
FT /function= "promoter plac"
FT /evidence= EXPERIMENTAL
FT -35_signal 23..28
FT /tag= b
FT /standard_name= "promoter plac"
FT 45..50
FT /tag= c
FT /evidence= EXPERIMENTAL
FT -10_signal 105..109
FT /tag= d
FT /evidence= EXPERIMENTAL
FT RBS /standard_name= "bioB RBS no. 9"
FT 117..1157
FT /tag= e
FT /product= "biotin synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioB"
FT /number= 1
FT 1141..1146
FT /tag= f
FT /standard_name= "bioF RBS"
FT 1154..2311
FT /tag= g
FT /EC_number= 2.3.1.47
FT /product= "KAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioF"
FT /number= 2
FT /standard_name= "8-amino-7-oxononanoate synthase"
FT 2284..2288
FT /tag= h
FT /standard_name= "bioC RBS"
FT 2295..3050
FT /tag= i
FT /function= "involved in pimeloyl-CoA synthesis"
FT /product= "protein"
FT /gene= "bioC"
FT /number= 3
FT 3030..3033
FT /tag= j
FT /standard_name= "bioD RBS"
FT 3043..3753
FT /tag= k
FT /EC_number= 6.3.3.3
FT /product= "DPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioD15"
FT /number= 4
FT /standard_name= "delthiobiotin synthase"
FT 3712..3750
FT /tag= l
FT /note= "bioD15 substitution"
FT 3742..3746
FT /tag= m
FT /standard_name= "bioA RBS"
FT 3750..5039
FT /tag= n
FT /EC_number= 2.6.1.62
FT /product= "DAPA synthase"
FT /evidence= EXPERIMENTAL
FT /gene= "bioA"
FT /number= 5
FT /standard_name= "S-adenosyl-L-methionine: 8-amino-

```

```

FT RBS 5088..5093
FT /tag= o
FT /standard_name= "ORF1 RBS"
FT 5098..5574
FT CDS
FT /tag= p
FT /function= "unknown, involved in biotin synthesis"
FT /product= "protein"
FT /evidence= EXPERIMENTAL
FT /gene= "ORF1"
FT /number= 6
FT 5583..5644
FT /tag= q
FT /standard_name= "rho-independent transcriptional
FT stem_loop 5583..5605
FT /tag= r
FT WO9408023-A.
FT 14-APR-1994.
FT PD
FT 01-OCT-1993; 93WO-EP02688.
FT PF
FT 02-OCT-1992; 92CH-0003124.
FT PR
FT 15-JUL-1993; 93CH-0002134.
FT XX
FT (LONZ ) LONZA AG.
FT PA
FT Birch O, Brass J, Fuhrmann M, Shaw N;
FT PI
FT WPI: 1994-135587/16.
FT DR
FT P-PSDB: AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT PT
FT Biotechnological biotin prodn. using enterobacterial biotin-gene
FT - providing vitamin H in high yield
FT XX
FT PS
FT Claim 1: Fig 6, Page 47-55 and 60-65; 92pp; German.
FT CC
FT The sequence is derived from plasmid pB030A-15/9 contg. the
FT CC bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
FT CC of biotin, arranged in a transcription unit. Microorganisms
FT CC contg. these DNA fragments or plasmids may be used in the prodn.
FT CC of biotin. Biotin (vitamin H) may prevent seborrhoea, dermatitis,
FT CC loss of appetite and tiredness.
FT XX
FT SQ
FT Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
FT
FT Query Match 100.0%; Score 25; DB 15; Length 5872;
FT Best Local Similarity 100.0%; Pred. NO. 0.13;
FT Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT QY
FT 1 gcgcagcaggtgcatgcgcagcatt 25
FT |||||
FT Db 195 gcgcagcaggtgcatgcgcagcatt 219
FT
FT RESULT 5
FT AAF15143
FT ID AAF15143 standard; cDNA; 860 BP.
FT XX
FT AC AAF15143;
FT XX
FT DT 13-MAR-2001 (first entry)
FT XX
FT DE Trichoderma reesei EST SEQ ID NO:7666.
FT XX
FT KW Multiple gene expression; filamentous fungal cell; EST;
FT KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
FT KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
FT KW culture condition; environmental stress; spore morphogenesis;
FT KW metabolic pathway engineering; catabolic pathway engineering; ss.

```

XX OS Trichoderma reesei.
 XX PN WO200056762-A2.
 XX PD 28-SEP-2000.
 XX PF 22-MAR-2000; 2000WO-US07781.
 XX PR 22-MAR-1999; 99US-0273623.
 XX PA (NOVO) NOVO NORDISK BIOTECH INC.
 XX PA (NOVO) NOVO NORDISK AS.
 XX PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX DR WPI: 2000-594572/56.
 XX PT Monitoring differential expression of genes in filamentous fungal cells
 XX PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 XX PT substrate of expressed sequence tags -
 XX PS Claim 89; Page 3094; 3161pp; English.
 CC CC The present invention describes a method for monitoring differential
 CC CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC CC expression of the same genes in one or more second filamentous fungal
 CC CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC CC are used in the methods for monitoring differential expression of genes
 CC CC in a first filamentous fungal (FF) cell relative to expression of the
 CC CC same genes in one or more second filamentous fungal cells. Monitoring
 CC CC the global expression of genes from FF cells allows the production
 CC CC potential of the microorganisms to be improved. New genes may be
 CC CC discovered, possible functions of unknown open reading frames can be
 CC CC identified and gene copy number variation and stability can be
 CC CC monitored. The expression of genes can be used to study how FF cells
 CC CC adapt to changes in culture conditions, environmental stress, spore
 CC CC morphogenesis, recombination, metabolic or catabolic pathway
 CC CC engineering. Using ESTs provides several advantages over genomic or
 CC CC random cDNA clones including elimination of redundancy as one spot on an
 CC CC array equals one gene or open reading frame, and organization of the
 CC CC microarrays based on function of the gene products to facilitate
 CC CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC CC all specifically claimed in the present invention.
 XX XX
 SO Sequence 860 BP; 216 A; 249 C; 229 G; 141 T; 25 other;
 OY Query Match 73.6%; Score 18.4; DB 21; Length 860;
 OY Best Local Similarity 95.0%; Pred. No. 62;
 OY Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 3 gcagcaggtgcatgcagcagc 22
 OY ||||||||| |||||||
 OY Db 402 gcagcaggtgcatgcagcagc 421
 RESULT 6
 ID AAZ11737 standard; cDNA; 2533 BP.
 XX AAZ11737;
 XX AC AAZ11737;
 XX AC AAZ11737;
 XX DE 02-NOV-1999 (first entry)
 XX DE Human transport-associated protein-7 (TRANP-7) cDNA.
 XX DE Transport-associated protein; TRANP; nuclear pore; nuclear transport;
 XX DE vesicle trafficking; cancer; cystic fibrosis; multidrug resistance;
 KW

KW hypercholesterolaemia; diagnosis; treatment; ss.
 XX Homo sapiens.
 OS OS
 XX XX
 XX XX
 FH key Location/Qualifiers
 FT 36..2165
 FT CDS /*tag= a
 FT /product= "Human transport-associated protein-7"
 XX
 XX PN WO941373-A2.
 XX PD 19-AUG-1999.
 XX PF 05-FEB-1999; 99WO-US02527.
 XX PR 11-FEB-1998; 98US-0021764.
 XX PA (TNCY-) INCYTE PHARM INC.
 XX PI Au-Young J, Bandman O, Baughn MR, Corley NC, Guegler KJ;
 XX PI Hillman JL, Lal P, Yue H;
 XX DR WPI: 1999-508646/42.
 XX DR P-PSDB; AAY31645.
 XX PT Human TRANP coding sequences, used to treat transport disorders and
 XX PT cancer
 XX PS Claim 7; Page 82-83; 87pp; English.
 CC CC This sequence represents human transport-associated protein-7 (TRANP-7)
 CC CC cDNA. The DNA sequence was first identified in a human colon tissue
 CC CC cDNA library. The full-length cDNA was derived from a series of
 CC CC overlapping and/or extended cDNA sequences and is a consensus.
 CC CC TRANP-1 to 9 (AAY31639-Y31647) are a novel group of proteins with
 CC CC chemical and structural homology that are involved in molecular
 CC CC transport. Various disorders are associated with defects in the transport
 CC CC of molecules, either intracellularly or to the extracellular
 CC CC environment. Examples of such disorders include cystic fibrosis,
 CC CC multidrug resistance, hypercholesterolaemia and certain forms of diabetes
 CC CC mellitus. Defective nuclear transport may play a role in cancer. For
 CC CC example, the BRCA1 protein, associated with familial breast cancer, is
 CC CC aberrantly imported into the nucleus via nuclear pore complexes, but is
 CC CC normally located in the cytoplasm in breast cancer cells. In other
 CC CC cancers, cells can secrete excessive amounts of hormones e.g., cancers of
 CC CC the adrenal medulla can secrete excessive amounts of adrenaline and
 CC CC noradrenaline, leading to hypertension. TRANP is expressed in cancer
 CC CC cells, and transport disorders result from either excessive or
 CC CC insufficient molecular transport. Anti-TRANP antibodies and nucleic acids
 CC CC encoding TRANP can be used as diagnostic tools for such disorders. TRANP
 CC CC antagonists can be used to treat or prevent a cancer associated with
 CC CC increased TRANP expression. Anti-TRANP antibodies can be used directly
 CC CC as an antagonist or as a targeting mechanism for drugs. Alternatively,
 CC CC a TRANP antisense nucleotide can be used to treat cancers. A TRANP
 CC CC agonist or expression vector may be used to treat a disorder caused by
 CC CC reduced transport of biologically active molecules.
 XX XX
 SO Sequence 2533 BP; 610 A; 654 C; 756 G; 513 T; 0 other;
 OY Query Match 72.8%; Score 18.2; DB 20; Length 2533;
 OY Best Local Similarity 87.0%; Pred. No. 82;
 OY Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 1 gcgcagcaggtgcatgcagcagc 23
 OY ||||||||| |||||||||
 OY Db 1320 gcgcagcagcaggtgcatgcagcagc 1342
 RESULT 7
 ID AAF56388 standard; DNA; 2536 BP.
 XX AAF56388
 XX AAF56388

AC AAF56388;
 XX
 DT 23-APR-2001 (first entry)
 XX
 DE ABC transport related gene #1.
 XX
 KW Human: adenosine triphosphate-binding cassette;
 KW transporter proteins; ABC; immune; infection; cancer; neurological;
 KW cardiovascular; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200107658-A1.
 XX
 PD 01-FEB-2001.
 XX
 PE 20-JUL-2000; 2000MO-US19736.
 XX
 PR 23-JUL-1999; 99US-0145215.
 PR 18-AUG-1999; 99US-0149445.
 PR 12-NOV-1999; 99US-0164730.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Ni J, Moore PE;
 XX
 DR WPI; 2001-210950/21.
 XX
 PT New nucleic acid molecule encoding a human adenosine
 PT triphosphate-binding cassette transporter protein, useful in
 PT preventing, treating or ameliorating medical conditions such as cancer,
 PT Alzheimer's disease and microbial infections -
 XX
 PS Claim 1; Page 262-263; 260pp; English.
 XX
 CC The present invention relates to human adenosine
 CC triphosphate-binding cassette (ABC) transporter proteins.
 CC These proteins may be used to prevent, treat and diagnose
 CC diseases associated with inappropriate expression of ABC transport
 CC proteins. The disorders include for example: immune/autoimmune
 CC diseases (e.g. HIV (human immunodeficiency virus) infections,
 CC anemia, rheumatoid arthritis and multiple sclerosis), cancers
 CC and hyperproliferative disorders, neurological diseases (e.g.
 CC Alzheimer's disease, Parkinson's disease), cardio-/cerebrovascular
 CC disorders and infections caused by bacteria, viruses and fungi.
 CC
 SO Sequence 2536 BP; 611 A; 655 C; 756 G; 514 T; 0 other;

Query Match 72.8%; Score 18.2; DB 22; Length 2536;
 Best Local Similarity 87.0%; Pred. No. 82;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggcagcaggtgcatgcgcagca 23
 |||||
 DB 1323 ggcagcagcagcatgcgcagca 1345

RESULT 8
 AAA01491/c
 ID AAA01491 standard; cDNA: 300 BP.
 XX
 AC AAA01491;
 XX
 DT 19-MAY-2000 (first entry)
 XX
 DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:1482.
 XX
 KW Human: colon cancer; tumour; diagnosis; gene expression product;
 KW probe; detection; cancerous state; metastasis; identification;
 KW breast cancer; oestrogen receptor-positive breast cancer; therapy;
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.
 XX

OS Homo sapiens.
 XX
 PN W09958675-A2.
 XX
 PD 18-NOV-1999.
 XX
 PF 13-MAY-1999; 99MO-US10602.
 XX
 PR 14-MAY-1998; 98US-0085426.
 PR 15-MAY-1998; 98US-0085537.
 PR 15-MAY-1998; 98US-0085636.
 PR 21-OCT-1998; 98US-0105234.
 PR 27-OCT-1998; 98US-0105877.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSEQ) HYSEQ INC.
 XX
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
 XX
 DR WPI; 2000-126369/11.
 XX
 PT Polynucleotide library used to determine cancerous states of mammalian
 PT cells -
 XX
 PS Claim 1; Page 536; 1097pp; English.
 XX
 CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 CC libraries constructed from human colon cancer cell lines. The present
 CC invention also describes a method of detecting differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell, comprising
 CC detecting at least one differentially expressed gene product in a test
 CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived.
 CC The polynucleotide sequences can be used in a method for detecting
 CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer.
 CC
 SO Sequence 300 BP; 56 A; 89 C; 94 G; 61 T; 0 other;

Query Match 71.2%; Score 17.8; DB 21; Length 300;
 Best Local Similarity 90.5%; Pred. No. 1e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 ggcagcaggtgcatgcgcagca 23
 |||||
 DB 108 GCAGCAGGTGCTATCGCAGCA 88

RESULT 9
 AAA64352/c
 ID AAA64352 standard; cDNA: 1197 BP.
 XX
 AC AAA64352;
 XX
 DT 20-DEC-2000 (first entry)
 XX
 DE cDNA encoding a human G protein-coupled receptor designated GPR54.
 XX
 KW Human: GPR54; G protein-coupled receptor; GPCR; eating disorder;
 KW obesity; Alzheimer's disease; memory; pain; sexual disorder;
 KW growth hormone imbalance; gene therapy; ss.
 XX

		71.2%	Score 17.8;	DB 22;	Length 2048;
Query Match					
Best Local Similarity		90.5%;	Pred. No. 1.2e+02;		
Matches	19;	Conservative	0;	Mismatches	2;
				Indels	0;
				Gaps	0
QY	3 gcacgacgatgcattcgccacga	23			
DG	1596 GGAGCAGGTGCACCAAGCCACGA	1576			

inflammatory bowel disease, graft vs. host disease in bone marrow transplant, tuberculosis etc.. They may also be used for in vitro detoxification and prevention of endotoxin contamination. See AAM00468-81. Whole cell extracts from U373 cells were prepared to assess transcription factor NF-kappaB activation. Cells were seeded in 6-well plates at a density of 1 million cells per well one day prior to stimulation. For stimulation, purified soluble CD14 1-348, soluble CD14 delta(57-64) or soluble CD14(7-10)A was added at a final concentration of 20ng/ml with or without 20ng/ml of Re595 LPS for 20 hours. Protein concentration of the whole cell extracts was determined by micro BCA assay and ranged between 1.5-2 microg/microtitre. For examining the NF-kappaB complexes, electrophoretic mobility shift assays were performed. This sequence and its complement were annealed to generate a double stranded DNA containing the NF-kappaB binding site of the HIV virus long terminal repeat promoter. This annealed fragment was then filled with Klenow fragment and radioactively labelled and used as a probe.

Query Match	70.4%	Score 17.6;	DB 17;	Length 40;
Best Local Similarity	83.3%	Pred. No. 1e+02;		
Matches 20;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0
OY	1	gcgcagcaggtgcacgcccagcat 24		
Db	25	GGCTGCACGCTGCACGCCGGCAT 2		
	RESULT 13			
AAC03291				
ID	AAC03291	standard; cDNA; 384 BP.		
XX				
AC	AAC03291;			
XX				
DT	06-OCT-2000	(first entry)		
XX				
DE	Human secreted protein 5' EST, SEQ ID NO: 3289.			
XX				
KX	Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;			
XX	gene therapy; chromosome mapping; ss.			
XX				
OS	Homo sapiens.			
PN	EP1033401-A2.			
XX				
PD	06-SEP-2000.			
XX				
PF	21-FEB-2000; 2000EP-0200610.			
XX				
PR	26-FEB-1999; 99US-0122487.			
XX				
PA	(GEST) GENSET.			
XX				
PI	Dumas Milne Edwards J, Duclert A, Giordano J;			
XX				
DR	WPI; 2000-500381/45.			
XX				
DR	P-PSDB; AAG03285.			
XX				
PT	New nucleic acid that is a 5' expressed sequence tag (5' EST) for			
XX	obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for			
PT	diagnostic, forensic, gene therapy and chromosome mapping procedures -			
XX				
PS	Claim 1; SEQ ID 3289; 71pp + CD-ROM; English.			
XX				
XX	The present sequence is one of a large number of 5' ESTs derived from			
CC	mRNAs encoding secreted proteins. An ORF has been identified within the			
CC	sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs			
CC	derived from 30 different tissues. EST sequences usually correspond			
CC	mainly to the 3' untranslated region (UTR) of the mRNA because they are			
CC	often obtained from oligo-dT primed cDNA libraries. Such ESTs are not			
CC	well suited for isolating cDNA sequences derived from the 5' ends of			

CC mRNAs and even in those cases where longer CDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

XX Sequence 384 BP; 86 A; 107 C; 130 G; 61 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 384;
 Best Local Similarity 83.3%; Pred. No. 1.2e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 cgcagcagtgatcgcagcagcatt 25
 Db 181 cgcagctggtgcagccagcagcatt 204

RESULT 14

AAA60774/C

ID AAA60774 standard; cDNA, 1215 BP.

XX AAA60774;

DT 27-OCT-2000 (first entry)

DE Human SENR ligand nucleotide sequence SEQ ID NO:25.

XX SENR: sensory epithelium neurotensin-like receptor; urotensin II;
 KM diagnosis; G protein-coupled receptor; hypertension; GPR14; hormone;
 KM kidney disease; regulator; central function; circulatory function;
 KM heart function; immune system function; digestive function;
 KM metabolic function; genital function; ss.

XX Homo sapiens.

XX WO200032627-A1.

PN 08-JUN-2000.

PD 29-NOV-1999; 99WO-JP06649.

PF 30-NOV-1998; 98JP-0338984.

PR 04-FEB-1999; 99JP-0026848.

PR 26-AUG-1999; 99JP-0239367.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Mori M, Abe M, Shimomura Y, Sugo T, Kitada C;

XX WPI; 2000-412287/35.

DR Urotensin peptides which are ligands for sensory epithelium

PT neurotensin-like receptor (SENR) for diagnosis and treatment of

PT hypertension

XX Example 20; Page 137; 147pp; Japanese.

XX The present invention provides peptides which are ligands for sensory

XX epithelium neurotensin-like receptor (SENR), and their amides, esters

XX and salts. SENR is a G-protein coupled receptor protein (also known as

XX GPR14), and the peptides which are ligands for it are forms of the

XX peptide hormone urotensin II. The peptides can be used in the treatment

XX and diagnosis of hypertension and kidney disease, and the development of

XX drugs which are regulators of central functions, circulatory functions,

XX heart functions, immune system functions, digestive functions, metabolic

XX functions and genital functions. The present sequence represents a

XX human SENR ligand nucleotide sequence, which is used in an example from

XX the present invention.

XX Sequence 1215 BP; 141 A; 487 C; 393 G; 194 T; 0 other;

Query Match 70.4%; Score 17.6; DB 21; Length 1215;
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 ggcagcagtgatcgcagcagcatt 24
 Db 592 ggcagcagtgatcgcagcagcatt 569

RESULT 15

AAZ56381/C

ID AAZ56381 standard; DNA, 1344 BP.

XX AAZ56381;

DT 17-MAR-2000 (first entry)

DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:51.

XX Flagellin; flig; antigen; detection; ds.

XX Escherichia coli.

XX WO9961458-A1.

PN 02-DEC-1999.

PD 21-MAY-1999; 99WO-AU00385.

PR 21-MAY-1998; 98AU-0003634.

XX (UNSY) UNIT SYDNEY.

XX Reeves PR, Wang L;

XX WPI; 2000-072598/06.

DR Novel nucleic acid molecule useful for the detection of flagellated

PT bacterial strains in food, faeces, etc.

PS Claim 3; Page 225; 245pp; English.

CC AAZ56331 to AAZ56398 represent nucleic acid molecules (1) encoding all

CC or part of an Escherichia coli flagellin protein except a protein

CC expressed by E. coli H1, H7, H12 or H48 type strains. The present

CC invention also describes a method of detecting the presence of E. coli

CC of a particular H serotype in a sample, comprising specifically

CC hybridizing a nucleic acid, preferably at least a pair, derived from a

CC flagellating gene, specific for a particular flagellin gene associated

CC with the H serotype, to any E. coli in the sample which contain the gene,

CC and detecting any hybridised molecules, identifying the presence of that

CC serotype in the sample. (1) are useful for: (1) detecting the presence

CC of E. coli of H serotype in a sample by hybridising at least one or a

CC pair of (1) to any E. coli in the sample and detecting the hybridised

CC nucleic acid molecules; and (2) for detecting the presence of both O

CC and H-serotypes of E. coli by hybridising at least one or a pair of (1)

CC to any E. coli present in the sample and detecting the hybridised

CC nucleic acid molecules. (1) is particularly useful for detecting the

CC combination of O and H antigen. Hybridised (1) when using at least one

CC (1) is detected by southern blot analysis and, when using a pair of (1),

CC is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56420

CC represent primers used in the exemplification of the present invention.

XX Sequence 1344 BP; 362 A; 337 C; 322 G; 323 T; 0 other;

OY Query Match 70.4%; Score 17.6; DB 21; Length 1344;
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 cgcagcagtgatcgcagcagcatt 25

Db 600 CGTACCGGCGATCGCCAGCAT 577

RESULT 16
AAV4199/c
ID AAV41999 standard; DNA; 2126 BP.

XX AAV41999;

XX 20-NOV-1998 (first entry)

XX Human GPR14 (G-protein coupled receptor polypeptide) gene.

XX G-protein coupled receptor polypeptide; GPR14; human; fungal infection;

XX bacterial infection; protozoan infection; viral infection; agonist;

XX antagonist; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 694..1863

XX FT /*tag= a

XX FT /product= "human GPR14"

XX EP859052-A1.

XX 19-AUG-1998.

XX 18-NOV-1997; 97EP-0309252.

XX 27-JAN-1997; 97US-0789354.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Bergsma DJ, Shabon U;

XX WPI: 1998-429680/37.

XX P-PSDB: AAM59908.

XX New DNA sequence encoding G-protein coupled receptor polypeptide

XX GPR14 - and corresponding polypeptide, agonists, antagonists, etc.

XX Claim 3; Fig 1; 23pp; English.

XX This is the nucleotide sequence of the gene encoding the novel G-protein

XX coupled receptor polypeptide (GPR14), used in the method of the

XX invention. Human GPR14 polypeptides and polynucleotides can be used in

XX the treatment of infections such as bacterial, fungal, protozoan and

XX viral infections. Agonists and antagonists can be used to treat

XX conditions associated with Human GPR14 imbalance.

XX Sequence 2126 BP; 317 A; 763 C; 669 G; 377 T; 0 other;

XX Query Match 70.4%; Score 17.6; DB 19; Length 2126;

XX Best Local Similarity 83.3%; Pred. NO. 1.4e+02;

XX Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

XX 1 ggcagcagtgatcgccagcat 24

XX 1269 GCGCACGACCGCATGCCAGCAT 1246

XX RESULT 17

XX AA21174/c

XX ID AA21174 standard; CDNA; 2126 BP.

XX AA21174;

XX 04-NOV-1999 (first entry)

XX Human GPR14 coding sequence.

XX GPR14; G protein-coupled receptor; ischaemic coronary artery disease;

XX atherosclerosis; metabolic disease; CHF/myocardial dysfunction; migraine;

XX arrhythmia; restenosis; hypertension; hypotension; pulmonary disease;

XX fibrotic vasculopathy; cerebrovascular event; neurogenic inflammation;

XX haematopoietic disorder; adult respiratory distress syndrome; ARDS;

XX cancer; autoimmune disease; therapy; human; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 694..1863

XX FT /*tag= a

XX FT /product= GPR14

XX W09940192-A1.

XX 12-AUG-1999.

XX 27-JAN-1999; 99WO-US01634.

XX 15-JAN-1999; 99US-0232857.

XX 09-FEB-1998; 98US-0074075.

XX 10-APR-1998; 98US-0058725.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (SMIK) SMITHKLINE BEECHAM LAB PHARM.

XX (SMIK) SMITHKLINE BEECHAM PLC.

XX Alyar NV, Al-Barazanji K, Ames RS, Arnold AR, Bergsma DJ;

XX Chambers J, Douglas SA, Foley JJ, Gout B, Khandoudi N;

XX Sarau HM, Shabon U, Willette RN;

XX WPI: 1999-527305/44.

XX P-PSDB: AAY32920.

XX Human G protein-coupled receptor GPR14, useful for identifying

XX agonists and antagonists

XX Claim 4; Fig 1; 64pp; English.

XX This sequence encodes the human G protein-coupled receptor, GPR14, of the

XX invention. Human GPR14, polypeptide and polynucleotides are useful in

XX methods for treatment of, e.g. ischaemic coronary artery disease,

XX atherosclerosis, metabolic diseases, CHF/myocardial dysfunction,

XX arrhythmias, restenosis, hypertension and hypotension, pulmonary disease,

XX fibrotic vasculopathies, cerebrovascular events, neurogenic

XX inflammation/migraine, haematopoietic disorders, adult respiratory

XX distress syndrome (ARDS), cancer, autoimmune diseases, etc. The methods

XX can be used to identify agonists and antagonists of human and rat GPR14.

XX Sequence 2126 BP; 317 A; 763 C; 669 G; 377 T; 0 other;

XX Query Match 70.4%; Score 17.6; DB 20; Length 2126;

XX Best Local Similarity 83.3%; Pred. NO. 1.4e+02;

XX Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

XX 1 ggcagcagtgatcgccagcat 24

XX 1269 GCGCACGACCGCATGCCAGCAT 1246

XX RESULT 18

XX AAF61051/c

XX ID AAF61051 standard; DNA; 2214 BP.

XX AAF61051;

XX 16-MAY-2001 (first entry)

XX P. putida KT2440-associated DNA ORF07225.

XX DE

XX

Transgenic plant; detection; probe; amplification; vaccine carrier;
 microbial production strain; biological remediation; ds.
 Pseudomonas putida.
 DE19935088-A1.
 01-FEB-2001.
 27-JUL-1999; 99DE-1035088.
 27-JUL-1999; 99DE-1035088.
 (TIGR-) TIGR INST GENOMIC RES.
 (OUTA-) OUTAGEN GMBH.
 (GBR-) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
 (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
 (MED1-) MEDIZINISCHE HOCHSCHULE HANNOVER.
 WPI; 2001-192469/20.
 New DNA sequences specific for Pseudomonas putida KT2440, useful as
 safe genetic engineering host, allow detection in presence of other
 related bacteria.
 Claim 1a; Page 101-102; 158pp; German.
 This invention describes novel DNA sequences (1) for specific detection
 of Pseudomonas putida KT2440. The invention also describes (1)
 recombinant expression vector containing (1); (2) prokaryotic or
 eukaryotic cells transformed or transfected with (1) or the vector of
 (1); (3) production of expression products by culturing cells of (2);
 (4) expression products, or their fragments, of (1) and synthetic
 proteins or peptides with the same sequences (A); (5) poly- or
 mono-clonal antibodies (Ab) that react specifically with (A); (6)
 hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
 plants that contain transformed or transfected cells of (2); (8)
 carrying one or more (1), (1), and their fragments, are used as probes
 to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
 polymerase chain reaction, and for production of transgenic plants. (1),
 or antibodies that recognize their expression products, are used for
 detecting the presence of KT2440, particularly in presence of other,
 even closely related, bacteria. KT2440 is one of the bacteria classified
 as safe, by the National Institutes of Health, for genetic engineering
 work, e.g., as microbial production strains, for biological remediation
 and as vaccine carriers. (1) are exclusive to KT2440 with no significant
 homology with sequences in other bacteria (specifically the closely
 related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it
 has greater catabolic activity and better survival in, and adaptation to,
 the rhizosphere and soil.
 Sequence 2214 BP; 436 A; 713 C; 693 G; 372 T; 0 other;
 Query Match 70.4%; Score 17.6; DB 22; Length 2214;
 Best Local Similarity 83.3%; Pred. No. 1.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 gcgcagcagtgatgcagcagcat 24
 ||||| ||||| ||||| |||||
 Db 783 GCGCTGCGAGGTCGACGCCGCCAT 760
 RESULT 19
 AAT80740/C
 ID AAT80740 standard; DNA: 423 BP.
 AC AAT80740;
 XX
 DT 11-NOV-1997 (first entry)
 XX
 DE Type II topoisomerase database reference sequence SEQ ID NO:159.

Detection; identification; TopoII; contamination; food; human;
 animal; plant; soil; water; PCR; polymerase chain reaction; ds.
 Nocardia farcinica.
 US564594-A.
 08-JUL-1997.
 05-JUL-1990; 90US-0548138.
 06-JUN-1995; 95US-0470179.
 05-JUL-1990; 90US-0548138.
 13-AUG-1993; 93US-0106482.
 (UTAH) UNIV UTAH RES FOUND.
 Huang WM;
 WPI; 1997-362925/33.
 Detection and identification of organisms - using DNA primers to
 amplify signature segment of organism's type II topoisomerase
 Claim 6; Column 157-158; 114pp; English.
 A method has been produced for selectively amplifying DNA segments of
 one or more species of organisms in a sample. The method involves: (a)
 providing a database containing reference sequences, comprising a
 subunit sequence of a signature region of a macromolecule selected from
 a type II topoisomerase (TopoII), or a homologue, where each reference
 sequence is specific to a different species of a chosen group, and the
 macromolecule comprises 1st and 2nd conserved regions adjacently
 flanking the signature region; and (b) making an extract of DNA
 molecules, and selectively amplifying DNA segments of the signature
 region using a universal primer composition, comprising a primer
 constructed to bind a DNA encoding the macromolecule, to produce
 amplified DNA segments. The present sequence represents a DNA fragment
 containing a signature segment which is used in the database as a
 reference sequence for Nocardia farcinica. The method can be
 used to identify all of the different organisms present in a single
 sample without using multiple probes. It can accurately distinguish
 among similar and related species, and can be used with, e.g.,
 contaminated food products, tissue or fluid samples from diseased
 humans, animal or plants, soil samples and water samples from any
 source.
 Sequence 423 BP; 84 A; 144 C; 135 G; 60 T; 0 other;
 Query Match 68.8%; Score 17.2; DB 18; Length 423;
 Best Local Similarity 86.4%; Pred. No. 1.8e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 gcgcagcagtgatgcagcagc 22
 ||||| ||||| ||||| |||||
 Db 294 GCGCAGCATCTCCATCCGCGAGC 273
 RESULT 20
 AAF09857/C
 ID AAF09857 standard; cDNA: 568 BP.
 AC AAF09857;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Fusarium venenatum EST SEQ ID NO:2380.
 XX
 Multiple gene expression; filamentous fungal cell; EST;
 expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 Aspergillus oryzae; Trichoderma reesei; identification; recombination;

KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX Fusarium venenatum.
 OS
 XX WO200056762-A2.
 PN
 XX 28-SEP-2000.
 PD
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 XX 22-MAR-1999; 99US-0273623.
 PR
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 DR WPI; 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 86; Page 1260; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 SO Sequence 568 BP; 127 A; 155 C; 160 G; 114 T; 12 other;

Query Match 68.8%; Score 17.2; DB 21; Length 568;
 Best Local Similarity 86.4%; Pred. No. 1.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 gcagcagtgatgcgcagcat 24
 ||| ||||| ||||| |||
 DB 151 GCAGGAGTGCTCGCTCGCAGCGT 130

RESULT 21
 AAC58894/c
 ID AAC58894 standard; DNA; 769 BP.
 XX
 AC AAC58894;
 XX
 DT 25-JAN-2001 (first entry)
 XX
 DE Human tumour suppressor BRG1 gene exon 24.
 XX

KW Human; BRG1: tumour suppressor gene; cancer; chromosome 19p13.1;
 KW retinoblastoma tumour suppressor gene; RB: drug screening; gene therapy;
 KW drug design; peptide therapy; animal model; ss.
 XX Homo sapiens.
 OS
 XX WO200056931-A1.
 PN
 XX 28-SEP-2000.
 PD
 PF 23-MAR-2000; 2000WO-US07678.
 XX
 XX 23-MAR-1999; 99US-0125806.
 PR
 XX (MYRI-) MYRIAD GENETICS INC.
 PA
 XX Wong AKC, Tavtigian SV, Teng DH;
 PI WPI; 2000-587668/55.
 DR
 XX
 PT Diagnosing a polymorphism associated with predisposition for cancer in
 PT humans by determining whether there is a germline alteration of a BRG1
 PT gene or its expression products -
 XX
 PS Claim 18; Page 106; 215pp; English.
 XX
 CC The present invention is concerned with the use of the human tumour
 CC suppressor gene BRG1 in cancer diagnosis and therapy. This gene is
 CC comprised of several exons, shown in AAC58894-C58903, and has several
 CC splice variants, given in AAC58906-C58912. The protein sequences for
 CC these are shown in AAB27552-B27558. BRG1 is a homologue of the Drosophila
 CC protein brim, and has been shown to be bound to retinoblastoma tumour
 CC suppressor protein RB. The BRG1 coding sequence and protein can be used
 CC in the diagnosis and treatment of cancer (for example by gene therapy),
 CC particularly prostate cancer, to identify drugs useful in the treatment
 CC of cancer and in the production of animal models for cancer.
 XX
 SO Sequence 769 BP; 139 A; 212 C; 222 G; 196 T; 0 other;

Query Match 68.8%; Score 17.2; DB 21; Length 769;
 Best Local Similarity 86.4%; Pred. No. 1.9e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 gcagcagtgatgcgcagcat 24
 ||| ||||| ||||| |||
 DB 234 GCAGGAGTGCTCGCTCGCCGCT 213

RESULT 22
 AAQ31876/c
 ID AAQ31876 standard; DNA; 1462 BP.
 XX
 AC AAQ31876;
 XX
 DT 22-APR-1993 (first entry)
 XX
 DE Cyclin D2 pseudogene.
 XX
 KW Cyclin; D2; D3; lambdaD2-G1; lambdaD3-G5; destruction box; A; B;
 KW ubiquitin-dependant; degradation; similarity; D1; E; ss.
 XX
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 378..1327
 FT /*tag= a
 FT /tag= b
 FT /note= "Insertion"
 FT 495..497
 FT mutation
 FT /*tag= c
 FT /note= "Nonsense mutation"

```

FT mutation 956..958
FT /tag= d
FT /note= "Nonsense mutation"
FT misc_RNA 1002..1039
FT /tag= e
FT /note= "Deletion"
FT misc_RNA 1190
FT /tag= f
FT /note= "Inserted additional A residue"
FT misc_RNA 1290..1294
FT /tag= g
FT /note= "Encodes Ala or Ser"
FT mutation 1310..1312
FT /tag= h
FT /note= "Nonsense mutation"
XX
XX WO9220796-A.
XX
XX 26-NOV-1992.
XX
XX 18-MAY-1992; 92WO-US04146.
XX
XX 16-MAY-1991; 91US-0701514.
XX
XX (COLD-) COLD SPRING HARBOR LAB.
XX
XX Beach DH;
XX
XX WPI: 1992-415774/50.
XX
XX P-PSDB: AAR29313.
XX
XX Recombinant mammalian D-type cyclin - replaces a CLN-type protein
XX essential for cell start in budding yeast, its antipodes and
XX probes being useful in detecting D-type cyclin in biological
XX samples
XX
XX PS Disclosure; Fig 9; 75pp; English.
XX
XX CC The sequences given in AAQ31876-77 represents the cyclin D2 and D3
XX pseudogenes. These sequences were identified during the chromosomal
XX localisation of the D2 and D3 genes. These genes were contained
XX within clones lambdaD2-G1 and lambdaD3-G5 respectively. The three
XX novel D-type cyclins disclosed by this invention were shown to encode
XX a similar small size protein ranging from 289-295 amino acids. They
XX lack the "destruction box" identified in the N-terminus of both A-
XX and B-type cyclins which targets them for ubiquitin-dependant
XX degradation. This suggests that the D-type cyclins have evolved a
XX different mechanism to govern their periodic degradation during cell
XX cycle or that they do not undergo destruction. The human cyclin D
XX genes share very high similarity over their entire coding region: 60%
XX between D1 and D2, 60% between D2 and D3 and 52% between D1 and D3.
XX The cyclin D genes are more closely related than the cyclin B genes
XX but are less closely related than the cyclin A and cyclin E genes.
XX See also AAQ31873-75.
XX
XX SQ Sequence 1462 BP; 370 A; 388 C; 355 G; 349 T; 0 other;

Query Match 68.8%; Score 17.2; DB 13; Length 1462;
Best Local Similarity 86.4%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 ggcgcagcagtgatcgccagc 22
   ||||||| ||||| |||||
DB 392 GCCCAGCAGCTGCATGCCAGC 371

```

```

RESULT 23
AAQ53208/C
ID AAQ53208 standard; DNA; 1462 BP.
XX
AC AAQ53208;
XX

```

```

DT 22-JUN-1994 (first entry)
XX
XX Human cyclin D2 pseudogene.
DE
XX
XX D-type; mammalian; CLN protein; protein deficiency; cell cycle start;
XX yeast; complement; ds.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FT CDS 378..1327
FT /tag= a
XX
XX WO9324514-A.
XX
XX 09-DEC-1993.
XX
XX 25-MAY-1993; 93WO-US05000.
XX
XX 26-MAY-1992; 92US-0888178.
XX
XX (MITO-) MITOTIX.
XX
XX Beach DH;
XX
XX WPI: 1993-405720/50.
XX
XX P-PSDB: AAR44805.
XX
XX New D-type mammalian cyclin - replaces CLN-type protein needed
XX for cell start in budding yeast and is detected by antibodies or
XX hybridisation in biological samples to determine abnormal cell
XX division
XX
XX PS Disclosure; Fig 9; 108pp; English.
XX
XX CC The sequence is that of a human cyclin D2 pseudogene.
XX
XX SQ Sequence 1462 BP; 370 A; 388 C; 355 G; 349 T; 0 other;

Query Match 68.8%; Score 17.2; DB 14; Length 1462;
Best Local Similarity 86.4%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 1 ggcgcagcagtgatcgccagc 22
   ||||||| ||||| |||||
DB 392 GCCCAGCAGCTGCATGCCAGC 371

```

```

RESULT 24
AAT66495/C
ID AAT66495 standard; CDNA; 1815 BP.
XX
AC AAT66495;
XX

```

```

DT 24-NOV-1997 (first entry)
XX

```

```

DE Human glucose transporter GLUT-1 cDNA.
XX
XX Glucose transporter; GLUT-1; GLUT-2; chimeric transporter;
XX diabetes; gene therapy; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FT CDS 165..1643
FT /tag= a
XX

```

```

PN WO9715668-A2.
XX
PD 01-MAY-1997.
XX
PF 23-OCT-1996; 96WO-US17327.
XX

```

```

XX 23-OCT-1995; 95US-0546934.
XX (BETA-) BETAGENE INC.
XX (TEXA) UNIV TEXAS SYSTEM.
XX
XX Clark SA, Newgard CB, Normington KD, Thigpen AE;
XX WPI: 1997-297737/27.
XX P-PSDB; AAM17835.
XX
XX Use of glucose transporter, GLUT-2 or GLUT-2 chimera(s) for
XX cell-killing - used in negative and double selection protocols and
XX screening methods, for cancer treatment and treatment of diabetes
XX
XX Claim 138; Page 104-106; 169pp; English.
XX
XX This cDNA sequence codes for human glucose transporter GLUT-1
XX (AAM17835). A claimed polynucleotide comprises a contiguous nucleic
XX acid sequence from human GLUT-1 cDNA and rat GLUT-2 cDNA (see
XX AAT17835). It encodes a GLUT-1/GLUT-2 chimeric transporter that
XX confers glucose sensing capacity to a cell but which does not
XX render the cell subject to diabetic immune destruction, and which
XX does not transport streptozotocin. The claimed polynucleotide can
XX be administered to a patient to treat diabetes, or expressed in
XX a cell to prepare a recombinant cell that secretes insulin in
XX response to glucose and which can be administered to a patient to
XX treat diabetes.
XX
XX Sequence 1815 BP; 324 A; 572 C; 520 G; 399 T; 0 other;
XX
XX
XX Query Match 68.8%; Score 17.2; DB 18; Length 1815;
XX Best Local Similarity 86.4%; Pred. No. 2.1e+02;
XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 4 cagcaggtgcacgcagcatt 25
XX ||||||| ||||| |||||||
XX Db 470 CAGCAGGTTTCATCATCAGCATT 449
XX
XX RESULT 25
XX AA224632/c
XX ID AA224632 standard; cDNA; 2850 BP.
XX
XX AA224632;
XX
XX 07-DEC-1999 (first entry)
XX
XX Human lung tumor associated polynucleotide.
XX
XX Human: lung tumor; lung cancer; T cell stimulation; ss.
XX
XX Homo sapiens.
XX
XX WO9947674-A2.
XX
XX 23-SEP-1999.
XX
XX 17-MAR-1999; 99WO-US05798.
XX
XX 18-MAR-1998; 98US-0040802.
XX 18-MAR-1998; 98US-0040984.
XX 27-JUL-1998; 98US-0123912.
XX 27-JUL-1998; 98US-0123933.
XX
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Wang T;
XX WPI: 1999-571839/48.
XX
XX New isolated lung tumor polynucleotides, used to develop products for

```

```

PT the treatment, prevention and monitoring the progression of lung cancer
PT
XX Claim 12; Page 123-124; 148pp; English.
XX
XX The invention provides isolated human lung tumor nucleic acids and
XX polypeptides. The polypeptides can be used for the treatment of lung
XX cancer. The polypeptides and polynucleotides can be used to stimulate T
XX cells or antigen presenting cells for use in the treatment of lung
XX cancer. The polypeptides and monoclonal antibodies specific for the
XX polypeptides can also be used to inhibit the development of lung cancer.
XX Agents which bind the polypeptides can be used for detecting lung cancer
XX and for monitoring the progression of lung cancer.
XX
XX Sequence 2850 BP; 602 A; 800 C; 753 G; 695 T; 0 other;
XX
XX
XX Query Match 68.8%; Score 17.2; DB 20; Length 2850;
XX Best Local Similarity 86.4%; Pred. No. 2.2e+02;
XX Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
XX 4 cagcaggtgcacgcagcatt 25
XX ||||||| ||||| |||||||
XX Db 485 CAGCAGGTTTCATCATCAGCATT 464
XX
XX RESULT 26
XX AAC65871/c
XX ID AAC65871 standard; cDNA; 2856 BP.
XX
XX AAC65871;
XX
XX 21-FEB-2001 (first entry)
XX
XX Human lung cancer-associated cDNA for contig 33.
XX
XX Lung cancer; therapy; treatment; human; tumor; immunogenic; cytostatic;
XX vaccine; detection; ss.
XX
XX Homo sapiens.
XX
XX WO200061612-A2.
XX
XX 19-OCT-2000.
XX
XX 03-APR-2000; 2000WO-US08896.
XX
XX 02-APR-1999; 99US-0285479.
XX 17-DEC-1999; 99US-0466396.
XX 30-DEC-1999; 99US-0476496.
XX 10-JAN-2000; 2000US-0480884.
XX 22-FEB-2000; 2000US-0510376.
XX
XX (CORI-) CORIXA CORP.
XX
XX Wang T, Fan L;
XX WPI: 2000-628399/60.
XX
XX Isolated polypeptide comprising an immunogenic portion of a lung tumor
XX protein is used for detecting and monitoring progression of lung cancer
XX in a patient.
XX
XX Claim 25a; Page 156-157; 261pp; English.
XX
XX This invention describes a novel isolated polypeptide (I) which
XX comprising an immunogenic portion of a lung tumor protein or variant (P2)
XX which have cytostatic activity. The polypeptides and polynucleotides are
XX used in compositions and vaccines to inhibit the development of cancer,
XX especially lung cancer, in a patient. Methods described in the invention
XX can be used to monitor the progression of a cancer by carrying out the
XX detection at subsequent time points and comparing the results from the
XX different time points. CD4+ and/or CD8+ T-Cells isolated from a patient

```

CC are treated with P2, polynucleotides encoding P2 or antigen presenting
 CC cells expressing P2 and then administered to the patient to inhibit
 CC development of cancer.
 CC

SQ Sequence 2856 BP; 602 A; 804 C; 753 G; 697 T; 0 other;

Query Match 68.8%; Score 17.2; DB 21; Length 2856;
 Best Local Similarity 86.4%; Pred. No. 2,2e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 cagcaggtgcatcgccagcat 25
 ||||| |||| |||||
 Db 485 CAGCAGTTCATCATCAGCAT 464

RESULT 27
 AAA81538
 ID AAA81538 standard; DNA; 5532 BP.

AC AAA81538;

DT 04-DEC-2000 (first entry)

DE N. meningitidis partial DNA sequence gnm_85 SEQ ID NO:85.

KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;

KM antigen; vaccine; diagnosis; infection; antibacterial; identification;

XX Meningococcus B; MenB; ds.

OS Neisseria meningitidis.

PN W0200022430-A2.

PD 20-APR-2000.

PF 08-OCT-1999; 99WO-US23573.

PR 09-OCT-1998; 98US-0103794.

PR 30-APR-1999; 99US-0132068.

PA (CHIR) CHIRON CORP.

PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;

PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;

PI Rappuoli R, Pizza M;

XX

DR WPI: 2000-318079/27.

PT Isolated nucleotide sequences of Neisseria meningitidis which can be

PT used in the diagnosis and treatment of N. meningitidis infection and

PT other Neisserial infections, for example, N.gonorrhoea -

XX

PS Claim 7: Page 1480-1482; 1760pp: English.

CC The present invention describes methods of obtaining immunogenic

CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414

CC represent specifically claimed Neisseria meningitidis genomic DNA

CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent

CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to

CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the

CC isolation of Neisseria meningitidis DNA sequences; and AAA81254 to

CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF

CC sequences, which are all used in the exemplification of the present

CC invention. The nucleic acid sequences, protein sequences, and antibodies

CC against them, can be used in the manufacture of a composition. The

CC composition can be used as a medicament (or in the manufacture of a

CC medicament) for treating, preventing or diagnosing infection due to

CC Neisserial bacteria. For example, some of the identified proteins could

CC be components of vaccines against Meningococcus B; against all serotypes;

CC and/or against all pathogenic Neisseriae. Identification of sequences

CC from the bacterium will also facilitate production of biological probes,

CC particularly organism-specific probes. Attempts to make efficacious

CC Meningococcus B vaccines have failed mainly due to antigen tolerance.

CC Multivalent vaccines have also been tried but none have successfully

CC overcome antigenic variability. The provision of further, complete

CC sequences may provide an opportunity to identify secreted or surface

CC exposed proteins that may be presumed targets for the immune system and

CC which are not antigenically variable or at least more conserved than

CC other more variable regions.

XX

SQ Sequence 5532 BP; 1241 A; 1503 C; 1452 G; 1336 T; 0 other;

Query Match 68.8%; Score 17.2; DB 21; Length 5532;
 Best Local Similarity 86.4%; Pred. No. 2,3e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 gcagcaggtgcatcgccagcat 24
 ||||| |||| |||||
 Db 5390 gcagcaggtgcatcgccagcat 5411

RESULT 28
 AAF21611
 ID AAF21611 standard; DNA; 349980 BP.

AC AAF21611;

DT 13-MAR-2001 (first entry)

DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:112.

KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;

KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;

XX ds.

OS Neisseria meningitidis.

PN W0200066791-A1.

PD 09-NOV-2000.

PF 08-MAR-2000; 2000WO-US05928.

PR 30-APR-1999; 99US-0132068.

PR 08-OCT-1999; 99WO-US23573.

PR 28-FEB-2000; 2000GB-0004695.

PA (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX

XX Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignani V;

PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V,

PI Frazer CM, Grandi G;

XX

DR WPI: 2000-647603/62.

PT Neisseria meningitidis B full length genome sequence and open reading

PT frames are used to detect, treat and prevent Neisserial infections -

XX

PS Claim 7: Appendix A: 692pp: English.

CC The present invention describes the full length genome of

CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607

CC to AAF21613 represent fragments of the NMB genomic sequence, as the

CC sequence was too long to go in a record on its own it was split into 8

CC sequences which overlap each other at the beginning and end of each

CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at

CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at

CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the

CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to

CC AAF21606 represent PCR primers which are used in the exemplification of

CC the present invention. The NMB genome and fragments from it have

CC antibacterial activity, and can be used in vaccines and gene therapy.

CC Neisseria nucleic acids, proteins and/or antibodies which binds to the

Query Match 68.0%; Score 17; DB 21; Length 468;
Best Local Similarity 80.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 ggcgcagcagtgatgcagcagcatt 25
||||| ||||| ||| ||| |||
Db 44 ggcgcagcagtgatgcagcagcatt 68

RESULT 31
AAC49411/c
ID AAC49411 standard; DNA; 514 Bp.
XX
AC AAC49411;
XX
DT 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 61061.
DE
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.


```

XX (BIOT-) BIOTRANSPLANT INC.
PA Baetscher M.
XX
XX WPI: 1999-181034/15.
DR P-PSDB; AAM96323.
XX
XX Novel porcine stem cell factor variants - useful for generating
PT stem cells to support porcine bone marrow cell proliferation
XX
PS Example 1; Figure 1; 46pp; English.
XX
XX New versions of porcine stem cell factor (PSCF) which are membrane
CC bound but are active variants of the wild type protein. These
CC variants can be generated by amplifying regions of the coding
CC sequence and splicing these sequences together in recombinant
CC plasmid vectors which can then be used to transfect bone marrow
CC cells. The variant PSCF lacks exon 6. Preferably the variant is
CC generated as a splice variant but generally the variant is
CC synthesised to completely lack exon 6 or have exon 6 replaced with
CC a polynucleotide which encodes one or more amino acids (usually Gly
CC or poly-Gly). The variant PSCF protein and nucleic acids encoding
CC it can be used to generate stem cells for supporting proliferation
CC of pig bone marrow cells or other pluripotent or totipotent porcine
CC cells. This sequence is the full length wild-type PSCF.
XX
SQ Sequence 952 BP; 289 A; 183 C; 215 G; 265 T; 0 other;

Query Match          68.0%; Score 17; DB 20; Length 952;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggcgcagcaggtgcacgcgcagcatt 25
   | ||||| | |||| ||||| |||||
DB 716 gggcagcgcgtgcacgtgcgcagcatt 740

RESULT 36
AAX27708
ID AAX27708 standard; cDNA; 952 BP.
XX
XX AAX27708;
AC
XX 01-JUN-1999 (first entry)
DT
XX
XX Porcine stem cell factor (PSCF) encoding cDNA.
DE
XX
XX Porcine; totipotent cell; pluripotent; primordial germ cell; PGC;
KW porcine stem cell factor; transgenic pig; xenotransplantation; ES;
KW cell differentiation; gene regulation; embryonic development; PSCF;
KW embryonic stem cell; steel factor; ss.
XX
XX Sus sp.
OS
XX
XX W09909141-A1.
PN
XX
XX 25-FEB-1999.
PD
XX
XX 13-AUG-1998; 98WO-US16782.
PE
XX
XX 14-AUG-1997; 97US-0055643.
PR
XX
XX (BIOT-) BIOTRANSPLANT INC.
PA
XX
XX Baetscher M, Brem G;
PI
XX WPI: 1999-181024/15.
DR
XX P-PSDB; AAY01297.
PT
XX Production of pluripotent or totipotent porcine stem cell lines - by
PT long-term culture of transfected murine STO feeder cells with a

```

```

PT porcine stem cell factor, useful for, e.g. xenotransplantation
XX
XX Example 1; Fig 3A-B; 80pp; English.
PS
XX
XX The invention relates to an isolated porcine totipotent cell. A porcine
CC pluripotent or totipotent cell, can be produced by culturing either a
CC porcine primordial germ cell (PGC) or other totipotent cell with a
CC porcine stem cell factor (PSCF). Cell lines produced are useful for the
CC generation of transgenic pigs, and for xenotransplantation. They are also
CC useful for studying cell differentiation and gene regulation during
CC embryonic development. The use of totipotent or pluripotent cells, like
CC embryonic stem (ES) cells, in a totipotent-cell-embryo-injection method
CC enables specific gene alterations, which allow the study of specific gene
CC function in a resulting chimeric animal line. The present sequence
CC represents a cDNA encoding a PSCF (also termed as porcine "steel"
CC factor).
XX
XX
SQ Sequence 952 BP; 289 A; 183 C; 215 G; 265 T; 0 other;

Query Match          68.0%; Score 17; DB 20; Length 952;
Best Local Similarity 80.0%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 ggcgcagcaggtgcacgcgcagcatt 25
   | ||||| | |||| ||||| |||||
DB 716 gggcagcgcgtgcacgtgcgcagcatt 740

RESULT 37
AAF21879
ID AAF21879 standard; DNA; 2320 BP.
XX
XX AAF21879;
AC
XX
XX 27-MAR-2001 (first entry)
DT
XX
XX Human breast and ovarian cancer associated antigen gene SEQ ID 266.
DE
XX
XX Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;
KW neutrophic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; anticancer; vulnerary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiact; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW cardiovascular disorder; wound healing; neurological disease; ds.
XX
XX Homo sapiens.
OS
XX
XX W0200055173-A1.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 08-MAR-2000; 2000WO-US05881.
PE
XX
XX 12-MAR-1999; 99US-0124270.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX WPI: 2000-611515/58.
DR
XX P-PSDB; AAB58976.
PT
XX New human breast and ovarian cancer associated gene sequences and the
PT polypeptides encoded by these genes, useful in the prevention,
PT treatment and diagnosis of cancer, immune disorders, cardiovascular
PT disorders and neurological diseases -
PS
XX Claim 1; Page 685-686; 1299pp; English.
XX
XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human
CC

```



```

PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

```

```

PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match 68.0%; Score 17; DB 21; Length 2905;
Best Local Similarity 80.0%; Pred. No. 2.6e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

OY 1 ggcgcgcaggtgcacgcacatt 25
Db 370 GAGCAGCAGCAGCATTCACGCATT 346

```

```

RESULT 39
AAA55186
ID AAA55186 standard; DNA; 32998 BP.
XX AAA55186;

```

```

XX 30-AUG-2000 (first entry)
DE Cenarchaeum symbiosum nucleotide sequence variant A SEQ ID NO.1.
XX

```

```

XX Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;
KW characterisation; archae; therapeutic; industrial; laboratory; ds.
XX

```

```

XX Cenarchaeum symbiosum.
OS

```

```

XX WO200018909-A2.
PN

```

```

XX 06-APR-2000.
PD

```

```

XX 29-SEP-1999; 99WO-US22752.
PF

```

```

XX 29-SEP-1998; 98US-0102294.
PR

```

```

XX (DIVE-) DIVERSA CORP.
PA

```

```

XX Swanson RV, Feldman RA, Schleper C;
PI

```

```

XX WPI; 2000-293148/25.
DR

```

```

XX New nucleic acids and proteins isolated from the non-thermophilic
PT crenarchaeote Cenarchaeum symbiosum, useful in characterizing the
PT physiology of these archae and in therapeutic, industrial or laboratory
PT techniques -
PT

```

```

XX Claim 1; Page 64-74; 210pp; English.
PS

```

```

XX AAA55186 to AAA55226 and AAY90913 to AAY90951 represent nucleic acids
XX and proteins isolated from the non-thermophilic crenarchaeote
XX Cenarchaeum symbiosum. The nucleic acids and proteins identified in
XX the present invention are useful in characterizing the physiology of
XX these archae and can be used in therapeutic, industrial or laboratory
XX techniques. AAA55227 to AAA55260 represent promoter sequences from
XX Cenarchaeum symbiosum. AAA55261 to AAA55269 represent PCR primers and
XX probes used in examples from the present invention.

```

```

SO Sequence 32998 BP; 7312 A; 9457 C; 8851 G; 7378 T; 0 other;

```

```

Query Match 68.0%; Score 17; DB 21; Length 32998;

```

Key	Location/Qualifiers
ET CDS	417796..418671
ET	/*tag= a
ET	/standard_name= "ORF K1"

```

FT      /product= "oligopeptide permease"
FT      /note= "homologous to the oppc gene"
FT      418673..419680
FT      CDS
FT      /tag= b
FT      /standard_name= "ORF K2"
FT      /product= "oligopeptide permease"
FT      /note= "homologous to the oppd gene"
FT      419677..420738
FT      CDS
FT      /tag= c
FT      /standard_name= "ORF K3"
FT      /product= "oligopeptide permease"
FT      /note= "homologous to the oppf gene"
FT      420774..422159
FT      CDS
FT      /tag= d
FT      /standard_name= "ORF K4"
FT      /product= "encapsulation-like protein"
FT      /note= "homologous to the CapA gene"
FT      422628..424031
FT      CDS
FT      /tag= e
FT      /standard_name= "ORF K5"
FT      /product= "aminotransferase-like protein"
FT      /note= "homologous to the BioA gene"
FT      424056..425594
FT      CDS
FT      /tag= f
FT      /standard_name= "ORF K6"
FT      /product= "(semi)aldehyde dehydrogenase-like protein"
FT      /complement (426949..428028)
FT      CDS
FT      /tag= g
FT      /standard_name= "ORF K7"
FT      /product= "transposase homologue"
FT      /note= "homologous to the Tnp gene"
FT      428292..429623
FT      CDS
FT      /tag= h
FT      /standard_name= "ORF K8"
FT      /product= "glutamate dehydrogenase-like protein"
FT      /note= "homologous to the GUD1 gene"
FT      /complement (430538..431284)
FT      CDS
FT      /tag= i
FT      /standard_name= "ORF K9"
FT      /product= "transposase homologue"
FT      /complement (431296..432840)
FT      CDS
FT      /tag= j
FT      /standard_name= "ORF K10"
FT      /product= "transposase homologue"
FT      /note= "homologous to the Tnp gene"
FT      /complement (433880..434110)
FT      CDS
FT      /tag= k
FT      /standard_name= "ORF K11"
FT      /product= "protein of unknown function"
FT      /note= "homologous to the Fixu gene"
FT      /complement (434107..434433)
FT      CDS
FT      /tag= l
FT      /standard_name= "ORF K12"
FT      /product= "protein of unknown function"
FT      /complement (434517..434711)
FT      CDS
FT      /tag= m
FT      /standard_name= "ORF K13"
FT      /product= "ferrodoxin/ferrodoxin-like protein"
FT      /note= "homologous to the FdxN gene"
FT      /complement (434753..436234)
FT      CDS
FT      /tag= n
FT      /standard_name= "ORF K14"
FT      /gene= "nifB"
FT      /product= "protein involved in Femo co-factor biosynthesis"
FT      /complement (436460..438130)
FT      CDS
FT      /tag= o
FT      /standard_name= "ORF K15"
FT      /gene= "nifA"
FT      /product= "positive regulator of nif, fix and other genes"
FT      /complement (438297..438590)
FT      CDS
FT      /tag= p

FT      /standard_name= "ORF K16"
FT      /gene= "fixX"
FT      /product= "protein required for nitrogenase activity"
FT      /complement (438605..439912)
FT      CDS
FT      /tag= q
FT      /standard_name= "ORF K17"
FT      /gene= "fixC"
FT      /product= "protein required for nitrogenase activity"
FT      /complement (439923..441032)
FT      CDS
FT      /tag= r
FT      /standard_name= "ORF K18"
FT      /gene= "fixB"
FT      /product= "protein required for nitrogenase activity"
FT      /complement (441042..441899)
FT      CDS
FT      /tag= s
FT      /standard_name= "ORF K19"
FT      /gene= "fixA"
FT      /product= "protein required for nitrogenase activity"
FT      /complement (442316..442636)
FT      CDS
FT      /tag= t
FT      /standard_name= "ORF K20"
FT      /product= "protein of unknown function"
FT      /complement (443313..443879)
FT      CDS
FT      /tag= u
FT      /standard_name= "ORF K21"
FT      /product= "protein of unknown function"
FT      444337..445029
FT      CDS
FT      /tag= v
FT      /standard_name= "ORF K22"
FT      /product= "ferrodoxin-like protein"
FT      /note= "homologous to the Nifg gene"
FT      445088..446602
FT      CDS
FT      /tag= w
FT      /standard_name= "ORF K23"
FT      /gene= "dctA"
FT      /product= "C4-dicarboxylate transport protein"
FT      /note= "homologous to the DctAI gene"
FT      446599..447843
FT      CDS
FT      /tag= x
FT      /standard_name= "ORF L1"
FT      /product= "cytochrome P450-like protein"
FT      /note= "homologous to the Camc gene"
FT      447844..448500
FT      CDS
FT      /tag= y
FT      /standard_name= "ORF L2"
FT      /product= "gamma-hexachlorocyclohexane-dechlorinase-like protein"
FT      /note= "homologous to the LinA gene"
FT      448497..450203
FT      CDS
FT      /tag= z
FT      /standard_name= "ORF L3"
FT      /product= "putative protein with degradative function"
FT      450341..451396
FT      CDS
FT      /tag= aa
FT      /standard_name= "ORF L4"
FT      /product= "luciferase alpha-subunit-like protein"
FT      /note= "homologous to the LuxA gene"
FT      452980..454494
FT      CDS
FT      /tag= ab
FT      /standard_name= "ORF L6"
FT      /gene= "nifD"
FT      /product= "alpha-subunit of Femo protein of nitrogenase"
FT      454590..456131
FT      CDS
FT      /tag= ac
FT      /standard_name= "ORF L7"
FT      /gene= "nifK"
FT      /product= "beta-subunit of Femo protein of nitrogenase"
FT      456187..457677
FT      CDS
FT      /tag= ad
FT      /standard_name= "ORF L8"
FT      /product= "protein involved in Femo co-factor biosynthesis"
FT      /note= "homologous to the NifB gene"

```


CC (CTLA-4) sequence. The invention relates to a binding agent (I)
 CC comprising at least one monomeric VLD that is derived from a non-antibody
 CC ligand and has at least one CDR (complementarity determining region) loop
 CC sequence, or part of it, modified or replaced so that, compared to
 CC unmodified VLD, its solubility is increased and/or the size is altered
 CC and/or a disulphide bond is created within, or between, one or more CDR
 CC loops. (I) are used for diagnosis, e.g. in vivo detection/localisation of
 CC cancer, blood clots etc., also in vitro when immobilised on solid
 CC supports or biosensors and therapeutically. Modified VLD may have binding
 CC affinity for drugs, steroids, pesticides, antigens, growth factors,
 CC tumour markers, cell or viral proteins. Modification of VLD improves
 CC solubility and alters binding specificity. Since VLD are derived from
 CC human proteins, the need for a humanizing step (to avoid adverse immune
 CC responses) is avoided, and modification also improves expression in
 CC Escherichia coli.

SQ Sequence 68 BP; 10 A; 20 C; 23 G; 15 T; 0 other;

Query Match 66.4%; Score 16.6; DB 20; Length 68;
 Best Local Similarity 82.6%; Pred. No. 2.8e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcaggtgcacgcgcagcatt 25
 ||||| | ||||| ||||| ||
 Db 32 gcagccgagcagcgcgcagcatt 54

RESULT 45

AAZ20584
 ID AAZ20584 standard; DNA; 70 BP.

AC AAZ20584;

DT 23-NOV-1999 (first entry)

DE CTLA-4 variable like domain (VLD) specific primer.

XX CTLA-4: cytotoxic T-lymphocyte associated antigen 4; binding agent; VLD;

KM variable-like domain; human; diagnosis; cancer; blood clot; PCR primer;

XX ss.

OS Synthetic.

XX Homo sapiens.

PN MO9945110-A1.

XX 10-SEP-1999.

PD 05-MAR-1999; 99WO-AU00136.

XX 06-MAR-1998; 98AU-0002210.

PA (DIAP-) Diatech Pty Ltd.

PI Coia G, Galanis M, Hudson PJ, Irving RA, Nuttall SD;

XX WPI: 1999-551040/46.

XX New binding agent comprising monomeric V-like domain in which at least

PT one complementarity determining region loop is modified, useful for

XX diagnosis of cancer

PS Example 1; Fig 1; 117pp: English.

XX This sequence represents a PCR primer specific for the variable-like

CC domain (VLD) of the human cytotoxic T-lymphocyte associated antigen 4

CC (CTLA-4) sequence. The invention relates to a binding agent (I)

CC loops. (I) are used for diagnosis, e.g. in vivo detection/localisation of
 CC cancer, blood clots etc., also in vitro when immobilised on solid
 CC supports or biosensors and therapeutically. Modified VLD may have binding
 CC affinity for drugs, steroids, pesticides, antigens, growth factors,
 CC tumour markers, cell or viral proteins. Modification of VLD improves
 CC solubility and alters binding specificity. Since VLD are derived from
 CC human proteins, the need for a humanizing step (to avoid adverse immune
 CC responses) is avoided, and modification also improves expression in
 CC Escherichia coli.

SQ Sequence 70 BP; 10 A; 19 C; 24 G; 17 T; 0 other;

Query Match 66.4%; Score 16.6; DB 20; Length 70;
 Best Local Similarity 82.6%; Pred. No. 2.8e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcaggtgcacgcgcagcatt 25
 ||||| | ||||| ||||| ||
 Db 35 gcagccgagcagcgcgcagcatt 57

Search completed: October 9, 2001, 11:38:10
 Job time: 1805 sec

• •

•

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:39:41 ; Search time 221.13 Seconds
(without alignments)
21.403 Million cell updates/sec

Title: US-09-396-196f-5

Sequence: 1 gcgcagcagtgatcgcgcagcalt 25

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents.NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2 US-08-401-068-7	Sequence 7, Appl
2	25	100.0	1041	2 US-08-846-338-7	Sequence 7, Appl
3	25	100.0	5872	3 US-08-411-768B-1	Sequence 1, Appl
4	25	100.0	5872	3 US-08-411-768B-6	Sequence 6, Appl
5	17.6	70.4	410	2 US-08-484-397A-22	Sequence 22, Appl
6	17.6	70.4	2126	2 US-08-789-354-1	Sequence 1, Appl
7	17.6	70.4	2126	3 US-09-110-937-1	Sequence 1, Appl
8	17.6	70.4	2126	3 US-09-058-725B-1	Sequence 1, Appl
9	17.6	70.4	2126	3 US-09-232-857-1	Sequence 1, Appl
10	17.2	68.8	423	1 US-08-470-179-159	Sequence 159, App
11	17.2	68.8	1462	2 US-08-464-517-30	Sequence 30, Appl
12	17.2	68.8	1462	2 US-08-246-361A-30	Sequence 30, Appl
13	17.2	68.8	1462	3 US-08-463-772-30	Sequence 30, Appl
14	17.2	68.8	1462	5 PCT-US93-05000-30	Sequence 30, Appl
15	16.8	67.2	9551	1 US-08-056-2000-93	Sequence 93, Appl
16	16.8	67.2	9551	2 US-08-800-644-93	Sequence 93, Appl
17	16.6	66.4	414	3 US-08-804-180C-1	Sequence 1, Appl
18	16.6	66.4	423	1 US-08-470-179-45	Sequence 45, Appl
19	16.6	66.4	423	1 US-08-470-179-47	Sequence 47, Appl
20	16.6	66.4	484	6 5496550-5	Patent No. 5496550
21	16.6	66.4	561	1 US-08-067-684-13	Sequence 13, Appl
22	16.6	66.4	561	1 US-08-008-898-13	Sequence 13, Appl
23	16.6	66.4	561	2 US-08-459-818-13	Sequence 13, Appl
24	16.6	66.4	561	2 US-08-888-666-13	Sequence 13, Appl
25	16.6	66.4	561	2 US-08-465-078-13	Sequence 13, Appl
26	16.6	66.4	561	2 US-08-725-776-13	Sequence 13, Appl
27	16.6	66.4	561	2 US-08-488-062-13	Sequence 13, Appl

28	16.6	66.4	561	3 US-08-228-208A-13	Sequence 13, Appl
29	16.6	66.4	561	5 PCT-US95-06726-35	Sequence 35, Appl
30	16.6	66.4	1289	2 US-08-344-833-1	Sequence 1, Appl
31	16.6	66.4	2188	1 US-07-706-872-2	Sequence 2, Appl
32	16.6	66.4	3223	2 US-08-620-694A-9	Sequence 9, Appl
33	16.6	66.4	3223	3 US-08-022-255-9	Sequence 9, Appl
34	16.6	66.4	3223	3 US-09-022-696-9	Sequence 9, Appl
35	16.6	66.4	3223	3 US-08-978-773-3	Sequence 3, Appl
36	16.6	66.4	3223	3 US-09-022-253-9	Sequence 9, Appl
37	16.6	66.4	3223	3 US-09-022-260-9	Sequence 9, Appl
38	16.6	66.4	3223	4 US-09-022-259-9	Sequence 9, Appl
39	16.6	66.4	3223	4 US-09-022-257-9	Sequence 9, Appl
40	16.6	66.4	3958	1 US-08-435-933-5	Sequence 5, Appl
41	16.6	66.4	3958	5 PCT-US96-06035-5	Sequence 5, Appl
42	16.6	66.4	8931	3 PCT-US98-934-28	Sequence 28, Appl
43	16.2	64.8	117	6 5273901-8	Patent No. 5273901
44	16.2	64.8	117	6 5482709-7	Patent No. 5482709
45	16.2	64.8	423	1 US-08-470-179-144	Sequence 144, App

ALIGNMENTS

RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="biotin synthase"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcatccagcatt 25
|||||
DB 79 GCGCAGCAGTGCATCGCCAGCATT 103

RESULT 2

US-08-846-338-7
Sequence 7, Application US/08846338
Patent No. 5869719
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/846.338
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:

NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgcatccagcatt 25
|||||
DB 79 GCGCAGCAGTGCATCGCCAGCATT 103

RESULT 3

US-08-411-768B-1
Sequence 1, Application US/08411768B
Patent No. 6083712

GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9

FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 117
OTHER INFORMATION: /product= "Biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 1
FEATURE:

NAME/KEY: CDS
LOCATION: 2295..3050
OTHER INFORMATION: /codon_start= 2295
OTHER INFORMATION: /function= "involved in pimeloyl-coA synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "bioc"
OTHER INFORMATION: /number= 3
FEATURE:

NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3750
OTHER INFORMATION: /EC_number= 2.6.1.62
OTHER INFORMATION: /product= "DAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioA"
OTHER INFORMATION: /number= 5
OTHER INFORMATION: /standard_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate aminotransf."
OTHER INFORMATION:
FEATURE:
NAME/KEY: CDS

```

LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter plac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /number= 1
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.04;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagagtgatgcgcagcatt 25
Db 195 ggcgcagagtgatgcgcagcatt 219

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
;

APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pB030A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTPB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "biob RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biob RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
```

FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.04;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcagtgatccgcagcat 25
|||||
DB 195 GCGCAGCAGTGATCGCCAGCAT 219

RESULT 5
US-08-484-397A-22/C
Sequence 22, Application US/08484397A
Patent No. 5869055
GENERAL INFORMATION:

APPLICANT: Juan, Shao-chieh
APPLICANT: Lichenstein, Henri S.
TITLE OF INVENTION: ANTI-INFLAMMATORY CD14 POLYPEPTIDES
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMGEN INC
STREET: 1840 DeHavilland Drive
CITY: Thousand Oaks
STATE: CA
COUNTRY: US
ZIP: 91320-1789

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,397A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Curry, Daniel R.
REGISTRATION NUMBER: 32,727
REFERENCE/DOCKET NUMBER: A-324A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 805/447-1000
TELEFAX: 805/447-1090
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-484-397A-22

Query Match 70.4%; Score 17.6; DB 2; Length 40;
Best Local Similarity 83.3%; Pred. No. 31;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcagtgatccgcagcat 24
|||||
DB 25 GCGCTGCAGCTGCACCGCGGCAT 2

RESULT 6
US-08-789-354-1/C
Sequence 1, Application US/08789354
Patent No. 5851798
GENERAL INFORMATION:
APPLICANT: Shabon, Usman
APPLICANT: Bergsma, Derk

TITLE OF INVENTION: Cloning of Human GPR14 Re
TITLE OF INVENTION: Ceptor
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/789,354
FILING DATE: 27-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:
NAME: Han, William T
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: P50610
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-789-354-1

Query Match 70.4%; Score 17.6; DB 2; Length 2126;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcagtgatccgcagcat 24
|||||
DB 1269 GCGCAGCAGCGCATGCGCAGCAT 1246

RESULT 7
US-09-110-937-1/C
Sequence 1, Application US/09110937A
Patent No. 6005074
GENERAL INFORMATION:
APPLICANT: SHABON, USMAN
APPLICANT: BERGSMA, DERK
TITLE OF INVENTION: CLONING OF HUMAN GPR14 RECEPTOR
FILE REFERENCE: P50610-1
CURRENT APPLICATION NUMBER: US/09/110,937A
CURRENT FILING DATE: 1998-07-06
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 2126
TYPE: DNA
ORGANISM: HOMO SAPIENS
US-09-110-937-1

Query Match 70.4%; Score 17.6; DB 3; Length 2126;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

OY      1  gccgagcaggtgcacgcgcagcat 24
        ||||| ||| |||| ||||| |||||
DB      1269  gccgaccacgccgcgcatggccagcat 1246

RESULT      8
US-09-058-725B-1/C
; Sequence 1, Application US/09058725B
; Patent No. 6133420
;
; GENERAL INFORMATION:
; APPLICANT: Ames, Robert
; APPLICANT: Satriu, Henry
; APPLICANT: Foley, James
; APPLICANT: Chamber, Jon
; TITLE OF INVENTION: A Method of Finding Antagonist
; TITLE OF INVENTION: and Antagonist To Human and Rat GPR 14
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/058,725B
; FILING DATE: April 10, 1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/789,354
; FILING DATE: 27-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Han, William T.
; REGISTRATION NUMBER: 34,344
; REFERENCE/DOCKET NUMBER: GP50005-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5219
; TELEFAX: 610-270-5090
;
; TELEX:
;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2126 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
;
US-09-058-725B-1

Query Match      70.4%; Score 17.6; DB 3; Length 2126;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1  gccgagcaggtgcacgcgcagcat 24
        ||||| ||| |||| ||||| |||||
DB      1269  gccgaccacgccgcgcatggccagcat 1246

RESULT      9
US-09-232-857-1/C
; Sequence 1, Application US/09232857
; Patent No. 6159700
;
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS, STEPHEN
; APPLICANT: WILLEFFE, ROBERT
; APPLICANT: AIYAR, NAMBI
; APPLICANT: ROMANTIC, ANNE
; APPLICANT: KHANDOODJI, NASSIRAH
;

```

```

APPLICANT: GOOT, BERNARD
APPLICANT: AL-BARAZANJI, KAWAL
APPLICANT: AMES, ROBERT S.
APPLICANT: FOLEY, JAMES J.
APPLICANT: SARAU, HENRY
APPLICANT: CHAMBERS, JON K.
APPLICANT: SHABON, USMAN
APPLICANT: BERGSMAN, DEK
TITLE OF INVENTION: A METHOD OF FINDING AGONIST
AND ANTAGONIST TO HUMAN AND RAT GPR14
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ratner & Prestlia
STREET: P.O. Box 980
CITY: Valley Forge
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/232,857
FILING DATE: 15-JAN-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/789,354
FILING DATE: 27-JAN-1997
APPLICATION NUMBER: 60/074,075
FILING DATE: 09-FEB-1998
APPLICATION NUMBER: 09/058,725
FILING DATE: 10-APR-1998
ATTORNEY/AGENT INFORMATION:
NAME: Prestlia, Paul F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GP-50005-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0700
TELEX: 846169
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-232-857-1

Query Match 70.4%; Score 17.6; DB 3; Length 2126;
Best Local Similarity 83.3%; Pred. No. 44;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0

QY 1 gccgcagcagtgatccgcacgcat 24
||||| ||| |||| |||||
Db 1269 gccgcacagccgcgcatgccacgcat 1246

RESULT 10
US-08-470-179-159/c
Sequence 159, Application US/08470179
Patent No. 5645994
GENERAL INFORMATION:
APPLICANT: Huang, Ph. D, Wai Mun
TITLE OF INVENTION: Method and Compositions for
IDENTIFICATION OF SPECIES IN A SAMPLE
NUMBER OF SEQUENCES: 207
CORRESPONDENCE ADDRESS:
ADDRESSEE: Trask, Brittt and Rossa
STREET: P.O. Box 2550

```

CITY: Salt Lake City
STATE: Utah
COUNTRY: USA
ZIP: 84110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,179
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Sweigert Ph.D, Susan E.
REGISTRATION NUMBER: 36,289
REFERENCE/DOCKET NUMBER: 2601
TELECOMMUNICATION INFORMATION:
TELEPHONE: 801-532-1922
TELEFAX: 801-531-9168
INFORMATION FOR SEQ ID NO: 159:
SEQUENCE CHARACTERISTICS:
LENGTH: 423 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: not relevant
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: No. 5645994ardia farcinica
US-08-470-179-159

Query Match 68.8%; Score 17.2; DB 1; Length 423;
Best Local Similarity 86.4%; Pred. No. 57;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 ggcgacgagtgatcgccagc 22
||||| | |||||
Db 294 GCCGACGATCTCCATGCCGACG 273

RESULT 11
US-08-464-517-30/c
Sequence 30, Application US/08464517
Patent No. 5869640
GENERAL INFORMATION:
APPLICANT: BEACH, David H.
TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/464,517
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/963,308
FILING DATE: 16-OCT-1992
APPLICATION NUMBER: US 07/888,178
FILING DATE: 26-MAY-1992
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/701,514
FILING DATE: 16-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Matthew P. Vincent
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MII-004C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 1462 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(378..569, 662..1000, 1040..1189,
LOCATION: 1191..1292, 1292..1324)
US-08-464-517-30

Query Match 68.8%; Score 17.2; DB 2; Length 1462;
Best Local Similarity 86.4%; Pred. No. 63;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 ggcgacgagtgatcgccagc 22
||||| | |||||
Db 392 GCCGACGATCTCCATGCCGACG 371

RESULT 12
US-08-246-361A-30/c
Sequence 30, Application US/08246361A
Patent No. 5998582
GENERAL INFORMATION:
APPLICANT: BEACH, David H.
TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/246,361A
FILING DATE: 19-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/963,308
FILING DATE: 16-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/888,178
FILING DATE: 26-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/701,514
FILING DATE: 16-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Matthew P. Vincent
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: MII-004C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 30:


```

? SEQUENCE CHARACTERISTICS:
? LENGTH: 1461 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
? NAME/KEY: CDS
? LOCATION: join(378..569, 662..1000, 1040..1189)
? LOCATION: 1191..1292, 1292..1324)
? OS-08-246-361A-30

```

Query March	68.8%;	Score 17.2;	DB 2;	Length 1462;
Best Local Similarity	86.4%;	Pred. No. 63;		
Matches 19;	Conservative	0;	Mismatches 3;	Indels 0;
				Gaps 0
OY	1	gcgcagcaggtgcacgcgcagc	22	
Db	392	gccccagcagcgtgcacgccccagc	371	

```

US-08-463-772-30/c
; Sequence 30, Application US/08463772
; Patent No. 606501
;
; GENERAL INFORMATION:
; APPLICANT: BEACH, David H.
; TITLE OF INVENTION: D-TYPE CYCLIN AND USES RELATED THERETO.
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,772
; FILING DATE:
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/963,308
; FILING DATE: 16-OCT-1992
; APPLICATION NUMBER: US 07/888,178
; FILING DATE: 26-MAY-1992
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/701,514
; FILING DATE: 16-MAY-1991
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Mathew P. Vincent
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: MII-004C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
;
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1462 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(378..569, 662..1000, 1040..1189,
; LOCATION: 1191..1292, 1292..1324)
;
US-08-463-772-30

```

Query Match	68.8%	Score 17.2	DB 3	Length 1462
Best Local Similarity	86.4%	Pred. No. 63		
Matches 19	Conservative 0	Mismatches 3	Indels 0	Gaps 0

Qy 1 gcgcagcagtgcatcgcagc 22
 |||||
 Db 392 GCCCAGCAGCTGCATGGCCAGC 371

```

14 RESULT
15 PCT-US93-05000-30/C
16 Sequence 30, Application PC/TUS9305000
17 GENERAL INFORMATION:
18 APPLICANT: MITOTIX
19 TITLE OF INVENTION: D-Type Cyclin and Uses Related Thereto
20 NUMBER OF SEQUENCES: 42
21 CORRESPONDENCE ADDRESS:
22 ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
23 STREET: Two Milltia Drive
24 CITY: Lexington
25 STATE: Massachusetts
26 COUNTRY: US
27 ZIP: 02173
28
29 COMPUTER READABLE FORM:
30 MEDIUM TYPE: Floppy disk
31 COMPUTER: IBM PC compatible
32 OPERATING SYSTEM: PC-DOS/MS-DOS
33 SOFTWARE: PatentIn Release #1.0, Version #1.25
34 CURRENT APPLICATION DATA:
35 APPLICATION NUMBER: PCT/US93/05000
36 FILING DATE: 19930525
37
38 CLASSIFICATION:
39 PRIOR APPLICATION DATA:
40 APPLICATION NUMBER: US/07/888,178
41 FILING DATE: 26-MAY-1992
42 ATTORNEY/AGENT INFORMATION:
43 NAME: Granahan, Patricia
44 REGISTRATION NUMBER: 32,227
45 REFERENCE/DOCKET NUMBER: CSHL91-02A
46 TELECOMMUNICATION INFORMATION:
47 TELEPHONE: 617-861-6240
48 TELEFAX: 616-861-9540
49
50 INFORMATION FOR SEQ ID NO: 30:
51 SEQUENCE CHARACTERISTICS:
52 LENGTH: 1462 base pairs
53 TYPE: NUCLEIC ACID
54 STRANDEDNESS: double
55 TOPOLOGY: linear
56
57 MOLECULE TYPE: DNA (genomic)
58
59 PCT-US93-05000-30

```

Query Match	68.8%;	Score 17.2;	DB 5;	Length 1462;
Best Local Similarity	86.4%;	Pred. No. 63;		
Matches 19; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

```

QY      1  gcgcagcaggtgcatgcgcagc  22
          || ||||| ||||| |||||
Db      392 GCCCAGCAGCTGCATGCCAGC  371

```

RESULT 15
US-08-056-200-93
; Sequence 93, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Stienfert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul

TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
METHOD OF INVENTION: Methods of Using Same
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/056,200
FILING DATE: 30-APR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fedrick, Michael F.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 9551 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1507..1644
FEATURE:
NAME/KEY: Intron
LOCATION: 1645..2511
FEATURE:
NAME/KEY: CDS
LOCATION: 2512..8070
US-08-056-200-93

Query Match 67.2%; Score 16.8; DB 1; Length 9551;
Best Local Similarity 90.0%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 cagcagtgatcgcagca 23
||||| ||||| ||||| |||||
DB 6805 CAGCAGCTGCACCGCAGCA 6824

RESULT 16
US-08-800-644-93
Sequence 93, Application US/08800644
Patent No. 5938752
GENERAL INFORMATION:
APPLICANT: Steinert, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-gyu
APPLICANT: Chung, Soo-il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
METHOD OF INVENTION: Methods of Using Same
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach

STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,644
FILING DATE: 14-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/056,200
FILING DATE: 30-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Fedrick, Michael F.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 9551 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1507..1644
FEATURE:
NAME/KEY: Intron
LOCATION: 1645..2511
FEATURE:
NAME/KEY: CDS
LOCATION: 2512..8070
US-08-800-644-93

Query Match 67.2%; Score 16.8; DB 2; Length 9551;
Best Local Similarity 90.0%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 cagcagtgatcgcagca 23
||||| ||||| ||||| |||||
DB 6805 CAGCAGCTGCACCGCAGCA 6824

RESULT 17
US-08-804-180C-1
Sequence 1, Application US/08804180C
Patent No. 6107056
GENERAL INFORMATION:
APPLICANT: Martin K. Oaks
TITLE OF INVENTION: SCLT-4 and Its Soluble Products
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Thomas M. Wozny
STREET: 100 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.50 inch Disk
COMPUTER: IBM
OPERATING SYSTEM: DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/804,180C
FILING DATE: February 20, 1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
  NAME: Thomas M. Wozny
  REGISTRATION NUMBER: 28,922
  REFERENCE/DOCKET NUMBER: 3284-00003
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (414) 271-7590
  TELEFAX: (414) 271-5770
INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 414
    TYPE: Nucleic acid
    STRANDEDNESS: Double stranded
    TOPOLOGY: linear
  MOLECULE TYPE: cDNA
  HYPOTHEetical: no
  ANTI-SENSE: no
  ORIGINAL SOURCE:
    ORGANISM: Homo sapien
    DEVELOPMENTAL STAGE: Adult
    TISSUE TYPE: lymphnode
  IMMEDIATE SOURCE:
    CLONE: Z04/hscTLA-4/PCR3
  POSITION IN GENOME:
    CHROMOSOME/SEGMENT: 2q33
  FEATURE:
    NAME/KEY: Human scTLA-4 gene
    IDENTIFICATION METHOD: found by experiment
    OTHER INFORMATION: Expresses B7 binding protein
US-08-804-180C-1

```

```

Query Match          66.4%  Score 16.6; DB 3; Length 414;
Best Local Similarity 82.6%  Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      3 gcagcaggtgcacgcagcagcatt 25
        |||||  |||||  |||||  |||||
DB      35 GCAGCCGAGCATGCCAGCATT 57

```

```

RESULT 18
US-08-470-179-45/c
Sequence 45, Application US/08470179
Patent No. 5645994
GENERAL INFORMATION:
  APPLICANT: Huang Ph.D, Mai Mun
  TITLE OF INVENTION: Method and Compositions for
  NUMBER OF SEQUENCES: 207
  CORRESPONDENCE ADDRESS:
    STREET: P.O. Box 2550
    CITY: Salt Lake City
    STATE: Utah
    COUNTRY: USA
  ZIP: 84110
  COMPUTER READABLE FORM:
    MEDIUM TYPE: floppy disk
    OPERATING SYSTEM: IBM PC compatible
    SOFTWARE: Patentin Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/470,179
    FILING DATE:
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Sweigert Ph.D, Susan E.
    REGISTRATION NUMBER: 36,289
    REFERENCE/DOCKET NUMBER: 2601
    TELECOMMUNICATION INFORMATION:

```

```

TELEPHONE: 801-532-1922
TELEFAX: 801-531-9168
INFORMATION FOR SEQ ID NO: 45:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 423 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: not relevant
  MOLECULE TYPE: other nucleic acid
  DESCRIPTION: /desc = "gyra gene segment"
  HYPOTHEtical: NO
  ANTI-SENSE: NO
  ORIGINAL SOURCE:
    ORGANISM: Borrelia coriaceae
US-08-470-179-45

```

```

Query Match          66.4%  Score 16.6; DB 1; Length 423;
Best Local Similarity 82.6%  Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      3 gcagcaggtgcacgcagcagcatt 25
        |||||  |||||  |||||  |||||
DB      242 GCAGCAGAGCATGCCAGCAAT 220

```

```

RESULT 19
US-08-470-179-47/c
Sequence 47, Application US/08470179
Patent No. 5645994
GENERAL INFORMATION:
  APPLICANT: Huang Ph.D, Mai Mun
  TITLE OF INVENTION: Method and Compositions for
  NUMBER OF SEQUENCES: 207
  CORRESPONDENCE ADDRESS:
    STREET: P.O. Box 2550
    CITY: Salt Lake City
    STATE: Utah
    COUNTRY: USA
  ZIP: 84110
  COMPUTER READABLE FORM:
    MEDIUM TYPE: floppy disk
    OPERATING SYSTEM: IBM PC compatible
    SOFTWARE: Patentin Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/470,179
    FILING DATE:
    CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Sweigert Ph.D, Susan E.
    REGISTRATION NUMBER: 36,289
    REFERENCE/DOCKET NUMBER: 2601
    TELECOMMUNICATION INFORMATION:
      TELEPHONE: 801-532-1922
      TELEFAX: 801-531-9168
  INFORMATION FOR SEQ ID NO: 47:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 423 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: not relevant
      MOLECULE TYPE: other nucleic acid
      DESCRIPTION: /desc = "gyra gene segment"
      HYPOTHEtical: NO
      ANTI-SENSE: NO
      ORIGINAL SOURCE:
        ORGANISM: Borrelia crociduræ
US-08-470-179-47

```

Query Match 66.4%; Score 16.6; DB 1; Length 423;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 gcacgagtgatcgcgcagcatt 25
||||| | ||||| | |
Db 242 GCACGAGGAGCATCGCATCAAT 220

RESULT 20
5496550-5/c
; Patent No. 5496550
; APPLICANT: WALLACH, MICHAEL; PUGATSCH, THEA; MENCHER, DAVID
; TITLE OF INVENTION: METHOD OF REDUCING THE OUTPUT OF EIMERIA
; OOCYSTS FROM A NEWBORN CHICK
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/108,763
; FILING DATE: 17-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 642,219
; FILING DATE: 16-JAN-1991
; APPLICATION NUMBER: 310,603
; FILING DATE: 14-FEB-1989
; APPLICATION NUMBER: 155,245
; FILING DATE: 12-FEB-1988
; APPLICATION NUMBER: 896,611
; FILING DATE: 14-AUG-1986
; SEQ ID NO: 5;
; LENGTH: 484
5496550-5

Query Match 66.4%; Score 16.6; DB 6; Length 484;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 gcacgagtgatcgcgcagcatt 25
||||| | ||||| | |
Db 172 GCACGAGGAGCATCGCATCAAT 150

RESULT 21
US-08-067-684-13
; Sequence 13, Application US/08067684
; Patent No. 5434131
; GENERAL INFORMATION:
; APPLICANT: Linsley, Peter S.
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Damle, Nitin K.
; APPLICANT: Brady, William
; TITLE OF INVENTION: CTLA4 RECEPTOR AND METHODS FOR ITS USE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 225 South Lake Avenue, Suite 900
; CITY: Pasadena
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/067,684
; FILING DATE: 26-MAY-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Adriano, Sarah B.
; REGISTRATION NUMBER: 34,470
; REFERENCE/DOCKET NUMBER: 7848-1

TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310/312-9900
; TELEFAX: 310/479-8340
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Homo sapiens
US-08-067-684-13

Query Match 66.4%; Score 16.6; DB 1; Length 561;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 gcacgagtgatcgcgcagcatt 25
||||| | ||||| | |
Db 38 GCACGAGGAGCATCGCATCTT 60

RESULT 22
US-08-008-898-13
; Sequence 13, Application US/08008898
; Patent No. 5770197
; GENERAL INFORMATION:
; APPLICANT: Linsley, Peter S.
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Damle, Nitin K.
; APPLICANT: Brady, William
; TITLE OF INVENTION: CTLA4 RECEPTOR AND METHODS FOR ITS USE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheldon & Mak
; STREET: 201 South Lake Avenue, Suite 800
; CITY: Pasadena
; STATE: California
; COUNTRY: United States
; ZIP: 91101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/008,898
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/723,617
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandel, Saralynn
; REGISTRATION NUMBER: 31,853
; REFERENCE/DOCKET NUMBER: 7848
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 796-4000
; TELEFAX: (818) 795-6321
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-008-896-13

Query Match 66.4%; Score 16.6; DB 1; Length 561;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 gcacgagtgatcgccagcatt 25
|||||
Db 38 GCAGCGAGCATCGCCAGCCTT 60

RESULT 23
US-08-459-818-13
; Sequence 13, Application US/08459818
; Patent No. 5851795
; GENERAL INFORMATION:
; APPLICANT: Linsley, Peter S.
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Damle, Nitin K.
; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 1150 Santa Monica Blvd., Suite 400
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FastSeq 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,818
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Adriano, Sarah B.
; REGISTRATION NUMBER: 34,470
; REFERENCE/DOCKET NUMBER: 30436.35US02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310-445-1140
; TELEFAX: 310-445-9031
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..561
; US-08-459-818-13

Query Match 66.4%; Score 16.6; DB 2; Length 561;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 gcacgagtgatcgccagcatt 25
|||||
Db 38 GCAGCGAGCATCGCCAGCCTT 60

RESULT 24

US-08-889-666-13
; Sequence 13, Application US/08889666
; Patent No. 588579
; GENERAL INFORMATION:
; APPLICANT: Linsley, Peter S.
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Damle, Nitin K.
; APPLICANT: Brady, William
; APPLICANT: Kiener, Peter A.
; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 1150 Santa Monica Blvd., Suite 400
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/889,666
; FILING DATE: 08-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/375390
; FILING DATE: 18-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Adriano, Sarah B.
; REGISTRATION NUMBER: 34,470
; REFERENCE/DOCKET NUMBER: 30436-35US01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310-445-1140
; TELEFAX: 310-445-9031
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..561
; US-08-889-666-13

Query Match 66.4%; Score 16.6; DB 2; Length 561;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 gcacgagtgatcgccagcatt 25
|||||
Db 38 GCAGCGAGCATCGCCAGCCTT 60

RESULT 25
US-08-465-078-13
; Sequence 13, Application US/08465078
; Patent No. 5885796
; GENERAL INFORMATION:
; APPLICANT: Linsley, Peter S.
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Damle, Nitin K.
; APPLICANT: Kiener, Peter A.
; APPLICANT: Brady, William
; TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:

ADDRESS: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,078
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-35US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-465-078-13

Query Match
Best Local Similarity 66.4%; Score 16.6; DB 2; Length 561;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcagtgatcgcgcagcatt 25
||||| |||||||||
Db 38 GCAGCCGAGGATCGCCAGCTTT 60

RESULT 26
US-08-725-776-13
Sequence 13, Application US/08725776
Patent No. 5968510
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Dame, Milt K.
APPLICANT: Brady, William
APPLICANT: Kiener, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,776

FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-35US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-725-776-13

Query Match
Best Local Similarity 66.4%; Score 16.6; DB 2; Length 561;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcagtgatcgcgcagcatt 25
||||| |||||||||
Db 38 GCAGCCGAGGATCGCCAGCTTT 60

RESULT 27
US-08-488-062-13
Sequence 13, Application US/08488062
Patent No. 5977318
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Dame, Milt K.
APPLICANT: Brady, William
APPLICANT: Kiener, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,062
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-35US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 13:

```
SEQUENCE CHARACTERISTICS:
LENGTH: 561 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..561
US-08-488-062-13

Query Match      66.4%; Score 16.6; DB 2; Length 561;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcagtgatcgccagcatt 25
    ||||| 1 ||||| ||||| ||||| ||
Db 38 GCAGCCGAGCATCGCCAGCTTT 60

RESULT 28
US-08-228-208A-13
; Sequence 13, Application US/08228208A
; Patent No. 6090914
; GENERAL INFORMATION:
; APPLICANT: Linsley, Peter S.
; APPLICANT: Ledbetter, Jeffrey A.
; APPLICANT: Damle, Nitin K.
; APPLICANT: Brady, William
; APPLICANT: Wallace, Philip M.
; TITLE OF INVENTION: CTLA4/CD28lg HYBRID FUSION
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 1150 Santa Monica Boulevard, Suite 400
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/228,208A
; FILING DATE: 15-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/008,898
; FILING DATE: 22-JAN-1993
; APPLICATION NUMBER: 07/723,617
; FILING DATE: 27-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Adriano, Sarah B
; REGISTRATION NUMBER: 34,470
; REFERENCE/DOCKET NUMBER: 30436-300S01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 310 445-1140
; TELEFAX: 310 445-9031
;
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..561
```

```
US-08-228-208A-13

Query Match      66.4%; Score 16.6; DB 3; Length 561;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcagtgatcgccagcatt 25
    ||||| 1 ||||| ||||| ||||| ||
Db 38 GCAGCCGAGCATCGCCAGCTTT 60

RESULT 29
PCT-US95-06726-35
; Sequence 35, Application PC/YUS9506726
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Ligands for Induction of Antigen Specific
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, suite 510
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06726
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/253,783
; FILING DATE: 03 JUNE 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: RPI-016PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..561
; PCT-US95-06726-35

Query Match      66.4%; Score 16.6; DB 5; Length 561;
Best Local Similarity 82.6%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 gcagcagtgatcgccagcatt 25
    ||||| 1 ||||| ||||| ||||| ||
Db 38 GCAGCCGAGCATCGCCAGCTTT 60

RESULT 30
US-08-344-833-1/C
; Sequence 1, Application US/08344833
; Patent No. 5874280
; GENERAL INFORMATION:
```

```
APPLICANT: Kell, G nther
TITLE OF INVENTION: Recombinant Bovine Herpesvirus
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/344,833
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Blackstone, William B.
REGISTRATION NUMBER: 29,772
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bovine herpesvirus
FEATURE:
NAME/KEY: CDS
LOCATION: 142..1095
OTHER INFORMATION: /product= "Protein"
OTHER INFORMATION: /standard_name= "ORF-1"
FEATURE:
NAME/KEY: misc_RNA
LOCATION: 1..3
OTHER INFORMATION: /function= "stop-codon giv"
FEATURE:
NAME/KEY: misc_RNA
LOCATION: 4..141
OTHER INFORMATION: /function= "untranslated"
OTHER INFORMATION: region"
US-08-344-833-1

Query Match 66.4%; Score 16.6; DB 2; Length 1289;
Best local Similarity 82.6%; Pred. No. 1.1e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ggcgcagcagtgatcgccagca 23
||||| ||| ||| |||
Db 1077 GCGCAGCAGCAGCAGCAGCA 1055

RESULT 31
US-07-706-872-2/c
Sequence 2, Application US/07706872
Patent No. 5237056
GENERAL INFORMATION:
APPLICANT: Fischbach, Gerald D.
TITLE OF INVENTION: Identification of a Protein Which
PROMOTES THE SYNTHESIS OF ACETYLCHOLINE RECEPTORS AND USES
TITLE OF INVENTION: Therefor
NUMBER OF SEQUENCES: 8
```

```
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/706,872
FILING DATE: 19910529
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: H091-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2188 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 172..972
US-07-706-872-2

Query Match 66.4%; Score 16.6; DB 1; Length 2188;
Best local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cgcgcagcagtgatcgccagcat 24
||||| ||| ||| ||| |||
Db 1941 CACGAGGAGCTGATCTCCAGCAT 1919

RESULT 32
US-08-620-694A-9/c
Sequence 9, Application US/08620694A
Patent No. 5869286
GENERAL INFORMATION:
APPLICANT: Yao, Zhengbin
APPLICANT: Fanslow, William
TITLE OF INVENTION: No. 5869286el Receptor That Blinds IL-17
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Power Macintosh
OPERATING SYSTEM: Apple operating System 7.5.5
SOFTWARE: Microsoft Word for Apple, Version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/620,694A
FILING DATE: 21 MARCH 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/538,765
```


RESULT 33
 US-09-022-255-9/c
 Sequence 9, Application US/09022255
 Patent No. 6072033
 GENERAL INFORMATION:
 APPLICANT: Yao, Zhengbin
 APPLICANT: Sparlgs, Melanie
 APPLICANT: Fanslow, William
 TITLE OF INVENTION: No. 6072033el Receptor That Binds IL-17
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Immunex Corporation
 STREET: 51 University Street
 CITY: Seattle
 STATE: WA
 COUNTRY: USA
 ZIP: 98101
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: Apple Power Macintosh
 OPERATING SYSTEM: Apple Operating System 7.5.5
 SOFTWARE: Microsoft Word for Apple, Version 6.0.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/022,255
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: USSN 08/620,694
 FILING DATE: 21 MARCH 1996
 APPLICATION NUMBER: USSN 08/538,765
 FILING DATE: 7 AUGUST 1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: USSN 08/410,535

RESULT 34
US-09-022-696-9/c
Sequence 9, Application US/09022696
Patent No. 6072037
GENERAL INFORMATION:
APPLICANT: Yao, Zhenglin
APPLICANT: Springs, Melanie
APPLICANT: Fanslow, William
TITLE OF INVENTION: No. 6072037el Receptor That Binds IL-17
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Power Macintosh
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: Microsoft Word for Apple, Version 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,696
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/620,694
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/410,535
FILING DATE: 23 MARCH 1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,695
REFERENCE/DOCKET NUMBER: 2617-B

ADDRESS: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Power Macintosh
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: Microsoft Word for Apple, Version 6.0.1.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/022,253
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/620,694
FILING DATE: 21-MARCH-1996
APPLICATION NUMBER: USSN 08/538,765
FILING DATE: 7 AUGUST 1995

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/410,535
FILING DATE: 23 MARCH 1995

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,695
REFERENCE/DOCKEN NUMBER: 2617-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
TELEFAX: (206)

INFORMATION FOR SEQ. ID NO.: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 3223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:

Query Match	66.4%;	Score 16.6;	DB 4;	Length 3223;
Best Local Similarity	82.6%;	Pred. No. 1.2e+02;		

Matches	19;	Conservative	0;	Mismatches	4;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

QY	1	g	c	g	c	a	g	t	g	c	a	t	c	g	c	a	c	a	23
Db	158	G	A	C	A	G	A	G	A	C	A	G	C	C	C	C	A	G	136

RESULT 39
HS-09-032-

```

1  Sequence 9, Application US/09022257
2  Patent No. 6197525
3  GENERAL INFORMATION:
4  APPLICANT: Yao, Zhengbin
5  APPLICANT: Spriggs, Melanie
6  APPLICANT: Ranslow, William
7  TITLE OF INVENTION: IL-17
8  NUMBER OF SEQUENCES: 10
9  CORRESPONDENCE ADDRESS:
10 ADDRESS: Immunex Corporation
11 STREET: 51 University Street
12 CITY: Seattle
13 STATE: WA
14 COUNTRY: USA
15 ZIP: 98101
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: Apple Power Macintosh
19 OPERATING SYSTEM: Apple Operating System 7.5.5
20 SOFTWARE: Microsoft Word for Apple, Version 6.0.1
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/09/022,257
23 FILING DATE:
24 CLASSIFICATION:
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: 08/620,694
27 FILING DATE:
28 CLASSIFICATION:
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: 08/620,694
31 FILING DATE:
32 CLASSIFICATION:
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: USSN 08/410,535
35 FILING DATE: 23 MARCH 1995
36 CLASSIFICATION:
37 ATTORNEY/AGENT INFORMATION:
38 NAME: Perkins, Patricia Anne
39 REGISTRATION NUMBER: 34,695
40 REFERENCE/DOCKET NUMBER: 2617-B
41 TELECOMMUNICATION INFORMATION:
42 TELEPHONE: (206)587-0430
43 TELEFAX: (206)
44 INFORMATION FOR SEQ ID NO: 9:
45 SEQUENCE CHARACTERISTICS:
46 LENGTH: 3223 base pairs
47 TYPE: nucleic acid
48 STRANDEDNESS: single
49 TOPOLOGY: linear
50 MOLECULE TYPE: cDNA to mRNA
51 HYPOTHETICAL: NO
52 ANTI-SENSE: NO
53 ORIGINAL SOURCE:
54 ORGANISM: Human
55 STRAIN: IL-17 R (hct1a8 receptor)
56 FEATURE:
57 NAME/KEY: CDS
58 LOCATION: 93..2693
59 US-09-022-257-9

```

Query Match	66.4%;	Score 16.6;	DB 4;	Length 3223;
Best Local Similarity	82.6%;	Pred. No. 1.2e+02;		
Matches 19;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0

```
QY      1 ggcgagcagtgcacatcgccagca 23  
        | ||||| | | | | |||||  
Db     158 GAGCAGCAGGAGCAGCCCCACGA 136
```

RESULT 40
US-08-435-933-5

Query Match	66.48;	Score 16.6;	DB 1;	Length 3958;
Best Local Similarity	82.68;	Pred. No. 1.2e+02;		
Matches 19;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;

```

QY      1  ggcgcagcagtgcatcgccagca 23
          |||||
Db      289  GCGCAGCAGCAGCGAGCCACGA 311

```

RESULT 41

PCT-US96-06035-5
Sequence 5, Application PC/TUS9606035
GENERAL INFORMATION:
APPLICANT: Cully, Doris F.
APPLICANT: Arena, Joseph P.
APPLICANT: Pareess, Philip S.
APPLICANT: Liu, Ken K.
TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE
TITLE OF INVENTION: CHANNELS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jody M. Glessner
STREET: 126 East Lincoln Avenue - P.O. BOX 2000-0907
CITY: Railway
STATE: New Jersey
COUNTRY: US

```

      ZIP: 07065-0907
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: PCT/US96/06035
      FILING DATE:
      CLASSIFICATION:
      ATTORNEY/AGENT INFORMATION:
      NAME: Glessner, Jody M.
      REGISTRATION NUMBER: 32,838
      REFERENCE/DOCKET NUMBER: 19264 PCT
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (908) 594-3046
      TELEFAX: (908) 594-4720
      INFORMATION FOR SEQ ID NO: 5:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 3958 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: cDNA
      PCT-US96-06035-5

Query Match          66.4%; Score 16.6; DB 5; Length 3958;
Best Local Similarity 82.6%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 ggcgcagcagtgctgcgcgcagca 23
      ||||||| || |||||||
Db      289 gccgcagcagcgcgcgcgcgcagca 311

RESULT 42
US-09-028-934-28/c
; Sequence 28, Application US/09028934
; Patent No. 6117670
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Lam, Steven T.
; APPLICANT: Hammer, Philip E.
; APPLICANT: van Pee, Karl-Heinz
; APPLICANT: Kirner, Sabine
; APPLICANT: Young, Thomas R.
; TITLE OF INVENTION: Pyrolytic Biosynthesis Genes and Uses
; TITLE OF INVENTION: thereof
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 6117670artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/028,934
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
```

```

      ATTORNEY/AGENT INFORMATION:
      NAME: Meigs, J. Timothy
      REGISTRATION NUMBER: 38,241
      REFERENCE/DOCKET NUMBER: GC1506/CIP7
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: 919-541-8587
      TELEFAX: 919-541-8689
      INFORMATION FOR SEQ ID NO: 28:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 8931 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
      MOLECULE TYPE: DNA (genomic)
      HYPOTHETICAL: NO
      ANTI-SENSE: NO
      ORIGINAL SOURCE:
      ORGANISM: Burkholderia cepacia
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 657..2267
      OTHER INFORMATION: /product= "Prna"
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 2270..3355
      OTHER INFORMATION: /product= "Prnb"
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 3421..5121
      OTHER INFORMATION: /product= "Prnc"
      FEATURE:
      NAME/KEY: CDS
      LOCATION: 5145..6266
      OTHER INFORMATION: /product= "Prnd"
      US-09-028-934-28

Query Match          66.4%; Score 16.6; DB 3; Length 8931;
Best Local Similarity 82.6%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1 ggcgcagcagtgctgcgcgcagca 23
      ||||||| || |||||||
Db      5010 gtgcagcagcttcctgcgcgcgca 4988

RESULT 43
5273901-8
; Patent No. 5273901
; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
; AUGUSTINE, PATRICIA G.; DANFORTH, HARRY D.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
; SPOROOITE 21.5 KB ANTIGEN, AC-6B
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/581,693
; FILING DATE: 12-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1988
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO: 8
; LENGTH: 117
; 5273901-8

Query Match          64.8%; Score 16.2; DB 6; Length 117;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 3 gcagcaggtgcatcgccagca 23
 ||||| ||| |||||
 Db 49 gcagcagcagcagccgagca 69

RESULT 44

5482709-7
 Patent No. 5482709
 APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
 SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
 AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
 TITLE OF INVENTION: DIMERIA ANTIGENIC COMPOSITION WHICH
 ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
 NUMBER OF SEQUENCES: 10
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/148,432
 FILING DATE: 08-NOV-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 581,693
 FILING DATE: 12-SEP-1990
 APPLICATION NUMBER: 215,162
 FILING DATE: 05-JUL-1989
 APPLICATION NUMBER: 746,520
 FILING DATE: 19-JUN-1985
 APPLICATION NUMBER: 627,811
 FILING DATE: 05-JUL-1984
 SEQ ID NO: 7:
 LENGTH: 117
 5482709-7

Query Match 64.8%; Score 16.2; DB 6; Length 117;
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 gcagcaggtgcatcgccagca 23
 ||||| ||| |||||
 Db 49 gcagcagcagcagccgagca 69

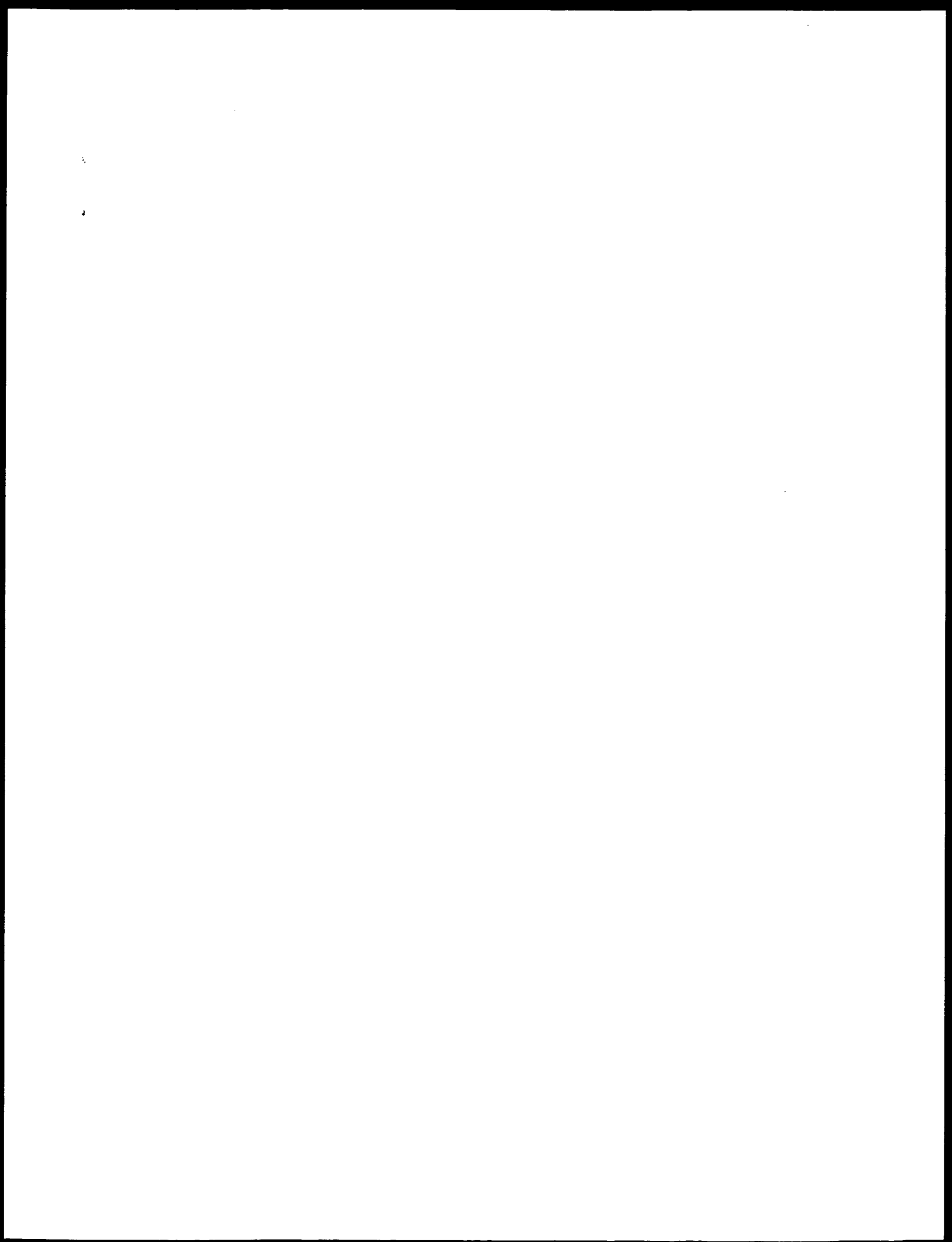
RESULT 45

US-08-470-179-144/c
 Sequence 144, Application US/08470179
 Patent No. 5645394
 GENERAL INFORMATION:
 APPLICANT: Huang Ph.D. Wei Mun
 TITLE OF INVENTION: Method and Compositions for
 IDENTIFICATION OF SPECIES IN A SAMPLE
 NUMBER OF SEQUENCES: 207
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Trask, Britt and Rossa
 STREET: P.O. Box 2550
 CITY: Salt Lake City
 STATE: Utah
 COUNTRY: USA
 ZIP: 84110
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/470,179
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Sweigert Ph.D. Susan E.
 REGISTRATION NUMBER: 36,289
 REFERENCE/DOCKET NUMBER: 2601
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 801-531-1922
 TELEFAX: 801-531-9168

INFORMATION FOR SEQ ID NO: 144:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 423 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: not relevant
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Mycobacterium fortuitum
 US-08-470-179-144

Query Match 64.8%; Score 16.2; DB 1; Length 423;
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 gcgcagcaggtgcatcgccag 21
 ||||| ||| |||||
 Db 294 GCGCAGCATCTCCATCGCCAG 274

Search completed: October 9, 2001, 11:39:43
 Job time: 1898 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 13:46:36 ; Search time 5930.9 Seconds
(without alignments)
39,846 Million cell updates/sec

Title: US-09-396-196f-5

Perfect score: 25

Sequence: 1 gcgcagcagtgatgcgcgcagcalt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_est3:*

4: gb_est4:*

5: gb_est5:*

6: gb_est6:*

7: gb_est7:*

8: gb_est8:*

9: gb_est9:*

10: gb_est10:*

11: gb_est11:*

12: gb_est12:*

13: gb_est13:*

14: gb_est14:*

15: gb_est15:*

16: gb_est16:*

17: gb_est17:*

18: gb_est18:*

19: gb_est19:*

20: gb_est20:*

21: gb_est21:*

22: gb_est22:*

23: gb_est23:*

24: gb_est24:*

25: gb_est25:*

26: gb_est26:*

27: gb_est27:*

28: gb_est28:*

29: gb_est29:*

30: gb_est30:*

31: gb_est31:*

32: gb_est32:*

33: gb_est33:*

34: gb_est34:*

35: gb_est35:*

36: gb_est36:*

37: gb_est37:*

38: gb_est38:*

39: gb_est39:*

40: gb_est40:*

41: gb_est41:*

42: gb_est42:*

43: gb_est43:*

44: gb_est44:*

45: gb_est45:*

46: gb_est46:*

47: gb_est47:*

48: gb_est48:*

49: gb_est49:*

50: gb_est50:*

51: gb_est51:*

52: gb_est52:*

53: gb_est53:*

54: gb_est54:*

55: gb_est55:*

56: gb_est56:*

57: gb_est57:*

58: gb_est58:*

59: gb_est59:*

60: gb_est60:*

61: gb_est61:*

62: gb_est62:*

63: gb_est63:*

64: gb_est64:*

65: gb_est65:*

66: gb_est66:*

67: gb_est67:*

68: gb_est68:*

69: gb_est69:*

70: gb_est70:*

71: gb_est71:*

72: gb_est72:*

73: gb_est73:*

74: gb_est74:*

75: gb_est75:*

76: gb_est76:*

77: gb_est77:*

78: gb_est78:*

79: gb_est79:*

80: gb_est80:*

81: gb_est81:*

82: gb_est82:*

83: gb_est83:*

84: gb_est84:*

85: gb_est85:*

86: gb_est86:*

87: gb_est87:*

88: gb_est88:*

89: gb_est89:*

90: gb_est90:*

91: gb_est91:*

92: gb_est92:*

93: gb_est93:*

94: gb_est94:*

95: gb_est95:*

96: gb_est96:*

97: gb_est97:*

98: gb_est98:*

99: gb_est99:*

100: gb_est100:*

101: gb_est101:*

102: gb_est102:*

103: gb_est103:*

104: gb_est104:*

105: gb_est105:*

106: gb_est106:*

107: gb_est107:*

108: gb_est108:*

109: gb_est109:*

110: gb_est110:*

111: gb_est111:*

112: gb_est112:*

113: gb_est113:*

114: gb_est114:*

115: gb_est115:*

116: gb_est116:*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est110:*
191: gb_est111:*
192: gb_hlc:*
193: em_gss_fun:*
194: em_gss_fun1:*
195: em_gss_fun2:*
196: em_gss_fun3:*
197: em_gss_fun4:*
198: em_gss_fun5:*
199: em_gss_fun6:*
200: em_gss_fun7:*
201: em_gss_fun8:*
202: em_gss_fun9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20.2	80.8	529	219	AG025369	AG025369 Oryza sat
C 2	19.8	79.2	506	175	BC266805	BC266805 1000101c1
C 3	19.8	79.2	527	258	P977L	AL630666 Leishman
C 4	19.8	79.2	582	23	A1665214	AL665214 605008C11
C 5	19.4	77.6	600	146	BF275487	BF275487 GA_Eb002
C 6	19.4	77.6	929	153	BG446019	BG446019 GA_Ea003
C 7	19.2	76.8	937	77	AW731362	AW731362 GA_Ea003
C 8	19.2	76.8	419	20	A1450725	A1450725 ms78c09.x
C 9	19.2	76.8	495	3	AA170657	AA170657 ms78c09.r
C 10	19.2	76.8	501	23	A1639758	A1639758 ms78c09.y
C 11	19.2	76.8	562	235	AQ922633	AQ922633 RPI-23-2
C 12	19.2	76.8	671	155	BG605917	BG605917 RH12-83-
C 13	19.2	76.8	707	145	BF183133	BF183133 601809516
C 14	18.8	75.2	323	107	AU093350	AU093350 AU093350
C 15	18.8	75.2	385	30	AV436765	AV436765 AV436765
C 16	18.8	75.2	475	12	AA821080	AA821080 GM09622.5
C 17	18.8	75.2	516	4	AA264635	AA264635 LD08220.5
C 18	18.8	75.2	541	258	TA80E010	TA80E010 T. brucei
C 19	18.8	75.2	637	231	AQ646542	AQ646542 RPI93-EC
C 20	18.8	75.2	781	153	BG417267	BG417267 HYSMEK001
C 21	18.6	74.4	499	256	B76887	B76887 T27F27F TAM
C 22	18.6	74.4	621	236	AQ969742	AQ969742 LERJUP67R
C 23	18.6	74.4	659	236	AQ969741	AQ969741 LERJUP67R
C 24	18.6	74.4	708	136	BE513502	BE513502 601315005
C 25	18.6	74.4	779	141	BE908400	BE908400 601503023
C 26	18.6	74.4	814	230	AQ576288	AQ576288 nxd00088N
C 27	18.6	74.4	1101	219	CNS0020N	AL061904 Drosophila
C 28	18.6	74.4	1332	141	BE871323	BE871323 601449431
C 29	18.4	73.6	343	30	AV397862	AV397862 AV397862
C 30	18.4	73.6	458	165	BE228836	BE228836 98A53192
C 31	18.4	73.6	786	221	CNS03HMV	AL244481 Tetradon
C 32	18.2	72.8	165	103	A1901905	A1901905 618012B05
C 33	18.2	72.8	290	11	AA741641	AA741641 LmtL39P3/
C 34	18.2	72.8	334	107	AU095285	AU095285 AU095285
C 35	18.2	72.8	349	169	BF753362	BF753362 QVO-CR058
C 36	18.2	72.8	349	175	BG268511	BG268511 1000201E1
C 37	18.2	72.8	355	167	BE397105	BE397105 601289536
C 38	18.2	72.8	405	122	AA952110	AA952110 EST364180
C 39	18.2	72.8	407	21	A1546214	A1546214 LD47729.5
C 40	18.2	72.8	420	6	AA387004	AA387004 VC20903.r
C 41	18.2	72.8	436	151	BF587149	BF587149 FM1-32-GO
C 42	18.2	72.8	460	172	BE994016	BE994016 MR0-GM013
C 43	18.2	72.8	470	165	BE268250	BE268250 601125693
C 44	18.2	72.8	489	21	A1542824	A1542824 SD09292.5
C 45	18.2	72.8	504	162	BE018236	BE018236 db77d06.y

ALIGNMENTS

RESULT 1
AG025369/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AG025369 529 bp DNA
Oryza sativa DNA, 3' flanking sequence of Tos17 insertion,
clone:T7115T, genomic survey sequence.
AG025369
AG025369.1 GI:7684033
GSS: GSS (genome survey sequence).
Oryza sativa (sub-species:japonica, strain:ND0016,
cultivar:Nipponbare) DNA, clone_11b:PCR product directly amplified
from rice genomic DNA clone:T7115T.
Oryza sativa
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
Oryza.
Miyao,A. and Hirochika,H.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Rice insertion mutants
Unpublished (1999)
2 (bases 1 to 529)
Miyao,A., Miyazaki,A., Yamashita,Y. and Hirochika,H.
Submitted (25-OCT-1999) to the DDBJ/EMBL/GenBank databases. Akio
Miyao, National Institute of Agrobiological Resources, Molecular
Genetics, 2-1-2, Kannondai, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:miyao@affrc.go.jp, URL:http://www.abr.affrc.go.jp/,
Tel:81-298-38-7006, Fax:81-298-38-7006)

FEATURES
source
Location/Qualifiers
1..529
/organism="Oryza sativa"
/cultivar="Nipponbare"
/strain="ND0016"
/sub-species="japonica"
/db_xref="taxon:4530"
/clone="T7115T"
/clone_11b="PCR product directly amplified from rice
genomic DNA"
/note="Sequence group name: T7115T. The 3' end of
retrotransposon Tos17 was found immediately upstream of
this sequence."

BASE COUNT 127 a 110 c 119 g 173 t
ORIGIN

Query Match 80.8%; Score 20.2; DB 219; Length 529;
Best Local Similarity 88.0%; Pred. No. 1.4e+02;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 1 gccgcagcaggtcgcgcagcatt 25
||||||| ||| |||||
Db 239 GCGCAGCAGCAGCATTGCCAGCATT 215

RESULT 2
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

BG266805 506 bp mRNA EST 20-FEB-2001
1000101G12.x1 1000 - Unigene 1 from Maize Genome Project Zea mays
cDNA, mRNA sequence.
BG266805
BG266805.1 GI:12970082
EST.
Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 506)
Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 1000101 row: G column: 12.
Location/Qualifiers
1..506
/organism="Zea mays"
/db_xref="dbEST:605008C11.x1"
/db_xref="taxon:4577"
/clone_11b="1000 - Unigene 1 from Maize Genome Project"
/note="This library represents the unique ESTs found in
the first round of EST sequencing at Stanford University
for the maize genome project. Sequences are present from
libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683
, 687, 707, and 945. Contigs were assembled using TIGR's

FEATURES High quality sequence stop: 584.
Location/Qualifiers
1. .600
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone_1lb="GA_Ea0024C21f"
/clone_1lb="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 170 a 94 c 160 g 176 t

ORIGIN

Query Match 77.6%; Score 19.4; DB 146; Length 600;
Best Local Similarity 95.2%; Pred. No. 3.1e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 gcagcagtgatcgccagca 23
|||||

DB 282 GCAGCAGTGCATCACCAGCA 262

RESULT 6
BG446019 929 bp mRNA EST 15-MAR-2001
LOCUS BG446019/c Gossypium arboreum 7-10 dpa fiber library Gossypium
DEFINITION arboeum cDNA clone GA_Ea0030G22f, mRNA sequence.
ACCESSION BG446019
VERSION BG446019.1 GI:13355671
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
1 (bases 1 to 929)
Wing, R.A., Frisch, D., Yu, Y., Main, D., Rambo, T., Simmons, J., Henry
D., Wood, T.C., Leslie, A. and Wilkins, T.A.
An integrated analysis of the genetics, development, and evolution
of the cotton fiber
Unpublished (2000)
JOURNAL COMMENT
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: TAAATACGACTCCTACTATAGG
High quality sequence start: 3
High quality sequence stop: 428.
Location/Qualifiers
1. .929
/organism="Gossypium arboreum"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone_1lb="GA_Ea0030G22f"
/clone_1lb="Gossypium arboreum 7-10 dpa fiber library"
/tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
/lab_host="E. coli"
/note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 259 a 114 c 367 g 187 t 2 others

ORIGIN

Query Match 77.6%; Score 19.4; DB 153; Length 929;
Best Local Similarity 95.2%; Pred. No. 3.2e+02;

Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 gcagcagtgatcgccagca 23
|||||

DB 290 GCAGCAGTGCATCACCAGCA 270

RESULT 7
AW731362/c standard; RNA; EST; 937 BP.
ID AW731362
XX
AC AW731362;
XX
SV AW731362.1
XX
DT 24-APR-2000 (Rel. 63, Created)
DT 18-NOV-2000 (Rel. 65, last updated, Version 2)
XX
DE GA_Ea0030G22 Gossypium arboreum 7-10 dpa fiber library Gossypium arboreum
DE CDNA clone GA_Ea0030G22, mRNA sequence.
XX
XX EST.
XX KW
OS Gossypium arboreum
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
XX
XX [1]
RP 1-937
RA Wing R.A., Frisch D., Yu Y., Main D., Rambo T., Simmons J., Henry D.,
RA Wood T.C., Leslie A., Wilkins T.A.;
RT "An integrated analysis of the genetics, development, and evolution of the
RT cotton fiber";
RL Unpublished.
XX
XX Contact: Wing RA
XX Clemson University Genomics Institute
XX Clemson University
XX 100 Jordan Hall, Clemson, SC 29634, USA
XX Tel: 864 656 7288
XX Fax: 864 656 4293
XX Email: twing@clemson.edu
XX High quality sequence stop: 937.
XX
XX Key Location/Qualifiers
FH 1. .937
FT source
FT /db_xref="taxon:29729"
FT /db_xref="ESTLIB:2480"
FT /note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"
FT /organism="Gossypium arboreum"
FT /strain="AKA"
FT /cultivar="8400"
FT /clone_1lb="GA_Ea0030G22"
FT /clone_1lb="Gossypium arboreum 7-10 dpa fiber library"
FT /tissue_type="Fibers isolated from bolls harvested 7-10 dpa"
FT /lab_host="E. coli"
FT
FT
SQ Sequence 937 BP; 261 A; 116 C; 371 G; 187 T; 2 other;

QY 3 gcagcagtgatcgccagca 23
|||||

DB 298 GCAGCAGTGCATCACCAGCA 278

RESULT 8

Query Match 77.6%; Score 19.4; DB 77; Length 937;
Best Local Similarity 95.2%; Pred. No. 3.2e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT	10
LOCUS	A1639758
DEFINITION	A1639758 501 bp mRNA EST 15-MAR-2000 ms7bc09.y1 Soares mouse 3MDMS Mus musculus cDNA clone IMAGE:617680
ACCESSION	A1639758
VERSION	A1639758.1 GI:4702867
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 501)
AUTHORS	Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person, B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurr,R., Ritter, E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,


```

QY      1  gcgcagcagtgatcgcacat  24
          |||||
Db     190  GCGCAGCAGCGGCCCGCCAGCAT  213

```

REFERENCE	1 (bases 1 to 385)
AUTHORS	Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and Tabata, S.
TITLE	Generation of 10,154 expressed sequence tags from a leaf gametophyte of a marine red alga, <i>Porphyra yezoensis</i>
JOURNAL	DNA Res. 7, 223-227 (2000)
MEDLINE	20363100

TITLE
JOURNAL

Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.

COMMENT

Direct Submission
Submitted (10-Dec-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at <http://www.sanger.ac.uk/projects/T-brucei/>.

FEATURES

source

Location/Qualifiers
1. .541
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="80e01"

BASE COUNT 113 a 141 c 151 g 136 t

ORIGIN

Query Match 75.2%; Score 18.8; DB 258; Length 541;
Best Local Similarity 90.9%; Pred. No. 5.6e+02;

Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY

1 ggcagcagtgatcgcgcacg 22
||||| ||||| ||||| |||||
Db 481 GCGCGCAGGAGCATCGCCAGC 502

RESULT 19
LOCUS

A0646542 637 bp DNA GSS 08-JUL-1999
DEFINITION RPI193-ECORI-6M17_TJ RPI193-ECORI Trypanosoma brucei genomic clone
ACCESSION A0646542
VERSION A0646542.1 GI:5123252

KEYWORDS

GSS.

SOURCE

Trypanosoma brucei.

ORGANISM

Trypanosoma brucei.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE

1 (bases 1 to 637)

AUTHORS

El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
Fraser, C. and Adams, M.

TITLE

Use of BAC end sequences from Trypanosoma brucei GUTat 10.1 RPI1-93
Library for gene discovery and sequence-ready map construction

JOURNAL

Unpublished (1999)

COMMENT

Other_GSS: RPI193-ECORI-6M17_TV
Contact: Najid M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones and high density filters may be purchased from BACPAC
Resources (<http://bacpac.med.buffalo.edu>). BAC end sequences search
page: <http://www.tigr.org/cdb/mdb/tbdb/>.
Seq primer: SP6
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1. .637
/organism="Trypanosoma brucei"

BASE COUNT

141 a 157 c 189 g 150 t

ORIGIN

Query Match 75.2%; Score 18.8; DB 231; Length 637;
Best Local Similarity 90.9%; Pred. No. 5.6e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ggcagcagtgatcgcgcacg 22
||||| ||||| ||||| |||||

Db 132 GCGCGCAGGAGCATCGCCAGC 153

RESULT 20
LOCUS

BGA17267 781 bp mRNA EST 13-MAR-2001
HVSMEK0017A16f Hordeum vulgare testa/pericarp EST library
HVCNDA0013 (normal) Hordeum vulgare cDNA clone HVSMEK0017A16f, mRNA
sequence.

ACCESSION BGA17267
VERSION BGA17267.1 GI:13322818

KEYWORDS

EST.

SOURCE

barley.

ORGANISM

Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae

REFERENCE

1 (bases 1 to 781)

AUTHORS

Wing, R., Close, T.J., Kleinjans, A., Wise, R., Begum, D., Frisch, D., Yu,
T., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo,
T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and
Wood, T.

TITLE

Development of a genetically and physically anchored EST resource
for barley genomics

JOURNAL

Unpublished (2000)

COMMENT

Contact: Ming RA

JOURNAL

Clemson University Genomics Institute

COMMENT

100 Jordan Hall, Clemson, SC 29634, USA

COMMENT

Tel: 864 656 7288

COMMENT

Fax: 864 656 4293

COMMENT

Email: rwing@clemson.edu

COMMENT

Seq primer: AATTACCTCCTCAAGAGG

COMMENT

High quality sequence stop: 753.

COMMENT

Location/Qualifiers

FEATURES

1. .781

source

/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEK0017A16f"
/clone_lib="Hordeum vulgare testa/pericarp EST library
HVCNDA0013 (normal)"
/issue_type="testa/pericarp"
/lab_host="TJ121"
/note="Vector: LambdaZAP, Site_1: EcoRI, Site_2: XhoI; For
more details on library preparation and sequence analysis
see <http://www.genome.clemson.edu/projects/barley/>"

BASE COUNT 191 a 211 c 253 g 126 t
 ORIGIN

Query Match 75.2%; Score 18.8; DB 153; Length 781;
 Best Local Similarity 90.9%; Pred. No. 5.7e+02;
 Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 4 cagcaggtcgtcgcagcatt 25
 |||||
 Db 291 CAGCAGCTGCATCAGCAGCT 312

RESULT 21
 B76887 499 bp DNA GSS 16-JAN-1998
 LOCUS B76887/c
 DEFINITION T27F2TF TAMU Arabidopsis thaliana genomic clone T27F2, DNA
 sequence.
 ACCESSION B76887
 VERSION B76887.1 GI:2773526
 KEYWORDS GSS.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 499)
 Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K.,
 Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter
 'J.C.
 A BAC End Sequence Database for Identifying Minimal Overlaps in
 Arabidopsis Genomic Sequencing. Update 3
 Unpublished (1997)
 Other GSSs: T27F2TR
 Contact: Steve Rounsley
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: rounsley@tigr.org
 Seq primer: M13-21
 Class: BAC ends
 High quality sequence stop: 499.
 Location/Qualifiers
 1..499
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="T27F2"
 /clone_1lb="TAMU"
 /sex="hermaphrodite"
 /note="Vector: BelobACIT; site_1: HindIII; site_2: HindIII
 ; Produced by Rod Wing"

BASE COUNT 141 a 90 c 116 g 152 t
 ORIGIN

Query Match 74.4%; Score 18.6; DB 256; Length 499;
 Best Local Similarity 84.0%; Pred. No. 6.8e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgcgcagcatt 25
 | ||||| ||||| ||||| |||||
 Db 218 GAGCAGCAGCGCGCATTCGACGATT 194

RESULT 22
 A0969742 621 bp DNA GSS 28-JAN-2000
 LOCUS A0969742/c
 DEFINITION LERJP67TR LERG Arabidopsis thaliana genomic clone LERJP67, DNA
 sequence.
 ACCESSION A0969742

VERSION A0969742.1 GI:6797443
 KEYWORDS GSS.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 621)
 Buell,C.R., Lin,X., Pal,G., Barnstead,M., Bowman,C., Utebach,T.,
 Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
 Genomic survey sequencing of Landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms

JOURNAL Unpublished (2000)
 COMMENT Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: atetigr.org
 For additional information, see <http://www.tigr.org/tdb/at/at.html>
 Seq primer: TR
 Class: Shotgun.

FEATURES
 source Location/Qualifiers
 1..621
 /organism="Arabidopsis thaliana"
 /strain="Landsberg erecta"
 /db_xref="taxon:3702"
 /clone="LERJP67"
 /clone_1lb="LERG"
 /note="Organ: Leaf; Vector: pUC19/K; Total genomic DNA was
 sheared to 0.4-0.7 Kbp before ligation."

BASE COUNT 177 a 108 c 125 g 209 t 2 others
 ORIGIN

Query Match 74.4%; Score 18.6; DB 236; Length 621;
 Best Local Similarity 84.0%; Pred. No. 6.8e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgcgcagcatt 25
 | ||||| ||||| ||||| |||||
 Db 525 GAGCAGCAGCTGCATTCGACGATT 501

RESULT 23
 A0969741 659 bp DNA GSS 28-JAN-2000
 LOCUS A0969741
 DEFINITION LERJP67TF LERG Arabidopsis thaliana genomic clone LERJP67, DNA
 sequence.
 ACCESSION A0969741
 VERSION A0969741.1 GI:6797442
 KEYWORDS GSS.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 659)
 Buell,C.R., Lin,X., Pal,G., Barnstead,M., Bowman,C., Utebach,T.,
 Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
 Genomic survey sequencing of Landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms

JOURNAL Unpublished (2000)
 COMMENT Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: atetigr.org
 For additional information, see <http://www.tigr.org/tdb/at/at.html>
 Seq primer: TF

TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: TATACGACTCACTATAGGCG
 Class: BAC ends
 High quality sequence stop: 427.
FEATURES
 source
 1..814
 /organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone_lib="CGI Rice BAC Library"
 /issue_type="Leaf"
 /lab_host="E. coli DH10B"
 /note="Vector: pBelBAC11, Site_1: HindIII; Site_2: HindIII. Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."
BASE COUNT 190 a 166 c 202 g 249 t 7 others
ORIGIN
 Query Match 74.4%; Score 18.6; DB 230; Length 814;
 Best Local Similarity 84.0%; Pred. No. 6.9e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ggcgcagcagtgatcgccacatt 25
 ||| ||||| ||||| ||||| |||||
Db 418 GTGCAGCAGCATTCGCCACATT 394
RESULT 27
LOCUS CENS0020N 1101 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BACR05E21 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL061904
KEYWORDS AL061904.1 GI:4940165
SOURCE GSS.
ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

COMMENT - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
 source
 1..1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="RPI-98"
 /clone="BACR05E21"
 /note="end : TET3"
BASE COUNT 275 a 202 c 223 g 319 t 82 others
ORIGIN
 Query Match 74.4%; Score 18.6; DB 219; Length 1101;
 Best Local Similarity 84.0%; Pred. No. 7e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 ggcgcagcagtgatcgccacatt 25
 ||| ||||| ||||| ||||| |||||
Db 819 GCCTACGACTGCACCGCACATT 795
RESULT 28
LOCUS BE871323 1332 bp mRNA EST 20-OCT-2000
DEFINITION 601449431f1 NIH_MGC_65 Homo sapiens CDNA IMAGE:3851614 5', mRNA sequence.
ACCESSION BE871323
KEYWORDS BE871323.1 GI:10320099
SOURCE EST.
ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1332)
 NIH-MGC <http://mgs.nci.nih.gov/>.
AUTHORS Mammalia: Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE NIH-MGC <http://mgs.nci.nih.gov/>.
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: LHAM9577 row: m column: 07
 High quality sequence stop: 381.
FEATURES
 source
 1..1332
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3851614"
 /clone_lib="NIH_MGC_65"
 /issue_type="adenocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: colon. Vector: pCMV-SPORT6. Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.8 kb. Library constructed by Life Technologies.
 BASE COUNT 317 a 544 c 276 g 195 t
 ORIGIN

Query Match 74.4%; Score 18.6; DB 141; Length 1332;
 Best Local Similarity 84.0%; Pred. No. 7e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 gccgacaggtgcatcgccacatt 25
 ||||| |||||
 Db 352 GCCCAGACATGCTTCCGACGACATT 376

RESULT 29
 AV397862/c

LOCUS AV397862 Chlamydomonas reinhardtii C9 Chlamydomonas reinhardtii
 DEFINITION CDNA clone CM002c09_r, mRNA sequence.
 ACCESSION AV397862.1 GI:6552078
 VERSION AV397862.1
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.
 1 (bases 1 to 343)
 Asamizu,E., Nakamura,Y., Sato,S., Fukuzawa,H. and Tabata,S.
 A large scale structural analysis of cDNAs in a unicellular green
 alga, Chlamydomonas reinhardtii. I. Generation of 343
 non-redundant expressed sequence tags
 DNA Res. 6 (6), 369-373 (1999)
 20152988

JOURNAL MEDLINE
 COMMENT Contact: Yasukazu Nakamura
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 source
 1..343
 /organism="Chlamydomonas reinhardtii"
 /strain="C9"
 /db_xref="taxon:3055"
 /clone="CM002c09_r"
 /clone_lib="Chlamydomonas reinhardtii C9"
 /dev_stage="photoautotrophic growth"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"
 BASE COUNT 62 a 108 c 101 g 72 t
 ORIGIN

Query Match 73.6%; Score 18.4; DB 30; Length 343;
 Best Local Similarity 95.0%; Pred. No. 8.1e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 gccgacaggtgcatcgcca 20
 ||||| |||||
 Db 100 GCCGACGACGCGCATCGCCA 81

RESULT 30
 BE228836/c
 LOCUS BE228836 458 bp mRNA
 DEFINITION 98AS3192 Rice Immature Seed Lambda ZAPII CDNA library Oryza sativa
 ACCESSION BE228836
 VERSION BE228836.1 GI:8955030
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

REFERENCE
 AUTHORS Nahn,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P.,
 Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee
 M.C. and Eun,M.Y.
 Large-scale Sequencing Analysis of ESTs from Rice Immature Seed
 Unpublished (1998)
 CONTACT: Eun M.Y.
 Department of Cytogenetics
 National Inst. of Agric. Sci. and Tech, RDA
 Suwon, Kyunggi-do, Korea
 Tel: 82 331 290 0301
 Fax: 82 331 290 0307
 Email: myeun@sun20.asi.re.kr.
 Location/Qualifiers
 1..458
 /organism="Oryza sativa"
 /cultivar="Milyang23"
 /db_xref="taxon:4530"
 /clone="98AS3192"
 /clone_lib="Rice Immature Seed Lambda ZAPII CDNA Library"
 /tissue_type="Immature Seed"
 /dev_stage="5 days after pollination"
 /lab_host="E. coli SDR"
 /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
 XhoI; Directional CDNA library inserted into lambda ZAPII
 vector at 5' end with EcoRI and 3' end with Xho I site."
 BASE COUNT 117 a 127 c 98 g 103 t
 ORIGIN

Query Match 73.6%; Score 18.4; DB 165; Length 458;
 Best Local Similarity 86.4%; Pred. No. 8.2e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 3 gccgacaggtgcatcgccacatt 24
 ||||| |||||
 Db 391 GCCGCGNGCATCGCAGCAT 370

RESULT 31
 CNS03HWM/c
 LOCUS CNS03HWM 786 bp DNA
 DEFINITION Tetradon nigroviridis genome survey sequence T7 end of clone
 027C16 of library G from Tetradon nigroviridis, genomic survey
 sequence.
 ACCESSION AL244481.1 GI:7965493
 VERSION AL244481
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetradon.
 1 (bases 1 to 786)
 Roest-Crollius H., Jallion O., Dasilva C., Fzames C., Fisher C.,
 Bouneau L., Billault A., Queller F., Saurin W., Bernot A. and
 Weissenbach J.
 Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetradon nigroviridis
 Unpublished
 2 (bases 1 to 786)
 Bernot A., Fzames C., Winker P., Brothier P., Queller F.,
 Saurin W. and Weissenbach J.
 Human gene number estimate provided by genome wide analysis using
 Tetradon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 786)
 Genoscope.
 Direct Submission

JOURNAL MEDLINE
 COMMENT Contact: Yasukazu Nakamura
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: ynakamu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

JOURNAL COMMENT Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases. This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES Location/Qualifiers

source 1..786
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone="027C16"
 /clone_lib="G"
 /note="genoscope sequence ID : COB027B808LP1-end : 17"
 BASE COUNT 236 a 188 c 199 g 160 t 3 others
 ORIGIN

Query Match 73.6%; Score 18.4; DB 221; Length 786;
 Best Local Similarity 95.0%; Pred. No. 8.3e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 agcaggtgcatgcgcagcat 24
 ||||| ||||| ||||| |||||
 Db 496 AGCAGATGCATGCAGCAT 477

RESULT 32
 AI901905 165 bp mRNA EST 27-JUL-1999
 LOCUS 618012B05.x1 618 - Inbred Tassel cDNA library Zea mays cDNA, mRNA
 DEFINITION sequence.

ACCESSION AI901905
 VERSION AI901905.1 GI:5608238
 KEYWORDS EST.

SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 165)
 Walbot, V.
 Maize ESTs from various cDNA libraries sequenced at Stanford University

JOURNAL COMMENT Unpublished (1999)
 Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 618012 row: B column: 05.

FEATURES Location/Qualifiers
 source 1..165
 /organism="Zea mays"
 /db_xref="taxon:4577"
 /cultivar="Ohio43"
 /clone_lib="618 - Inbred Tassel cDNA library"
 /clone_type="tassel"
 /dev_stage="tassel length from 0.1 to 2.5 cm"
 /lab_host="XLOLR"
 /note="Organ: tassel; Vector: pAD-GAL4-2.1 (Hybridzap);
 Inbred tassel library from Schmidt lab"

BASE COUNT 39 a 52 c 53 g 20 t 1 others
 ORIGIN

Query Match 72.8%; Score 18.2; DB 103; Length 165;
 Best Local Similarity 87.0%; Pred. No. 9.6e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggcagcaggtgcatgcgcagca 23
 ||||| ||||| ||||| |||||
 Db 130 GCGCAGCAGGTACAGCCACGCA 152

RESULT 33
 AA741641 290 bp mRNA EST 10-DEC-1998
 LOCUS Lmlv39p3/248A Leishmania major promastigote full length cDNA
 DEFINITION library from early logarithmic stage (day 3) Leishmania major cDNA
 clone 248A 5', mRNA sequence.

ACCESSION AA741641 GI:2781620
 VERSION AA741641.1
 KEYWORDS EST.
 SOURCE Leishmania major.
 ORGANISM Leishmania major.
 Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 290)
 Almeida, R., Sampaio, I., Schneider, H. and Blackwell, J.M.
 Analysis of Leishmania major promastigote library from different stages of development
 Unpublished (1998)
 Contact: Blackwell JM
 Cambridge Institute for Medical Research
 Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road,
 Cambridge CB2 2XY, UK
 Tel: 01223 336 143
 Fax: 01223 331 206
 Email: jmb37@cus.cam.ac.uk

PCR Primers
 FORWARD: GTAACGACGCGCCAGT
 BACKWARD: GGAACGCTATGACCATG
 Seq primer: AATTACCTCCTCACTAAGCG
 High quality sequence stop: 290.
 Location/Qualifiers
 source 1..290
 /organism="Leishmania major"
 /strain="LV39"
 /db_xref="taxon:5664"
 /clone="248A"
 /clone_lib="Leishmania major promastigote full length cDNA
 library from early logarithmic stage (day 3)"
 /cell_type="Promastigote"
 /note="Vector: Lambda Zap II; Site_1: XhoI; Site_2: NotI"

BASE COUNT 62 a 86 c 77 g 65 t
 ORIGIN

Query Match 72.8%; Score 18.2; DB 11; Length 290;
 Best Local Similarity 87.0%; Pred. No. 9.8e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ggcagcaggtgcatgcgcagca 23
 ||||| ||||| ||||| |||||
 Db 77 GCAACGACGCTATGCTACGCA 99

RESULT 34
 AU095285 334 bp mRNA EST 30-JUN-2000
 LOCUS AU095285/c Rice cDNA from immature leaf including apical meristem
 DEFINITION (under short day condition) Oryza sativa cDNA clone E61108, mRNA
 sequence.

ACCESSION AU095285
 VERSION AU095285.1 GI:8857967
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 334)
 Sasaki, T. and Yamamoto, K.
 Rice cDNA from immature leaf including apical meristem (2000)
 Unpublished (2000)
 JOURNAL


```

Db      269 GCGCAGCAGTACGCCAGCA 291
|||||
RESULT 37
LOCUS   BE397105      355 bp      mRNA      EST      21-JUN-2000
DEFINITION 601289536f1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:362001 5',
            mRNA sequence.
ACCESSION BE397105
VERSION   BE397105.1 GI:9342470
KEYWORDS EST.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 355)
            NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cga@bbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM291 row: 9 column: 10
            High quality sequence stop: 355.
FEATURES
            source
            1..355
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:362001"
            /clone_lib="NIH_MGC_8"
            /tissue_type="Burkitt lymphoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: lymph. Vector: pOTB7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5'
            adaptor: GGCACGAG(G). Size-selected >500bp for average
            insert size 1.8kb. Library constructed by Ling Hong in
            the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
```

```

Query Match      72.8%; Score 18.2; DB 167; Length 355;
Best Local Similarity 87.0%; Pred. No. 9.8e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 ggcagcaggtgcctgcgcagca 23
|||||
Db      112 GCGCAGCAGTATGCCAGCA 134

RESULT 39
LOCUS   A1546214      407 bp      mRNA      EST      22-MAR-1999
DEFINITION LD47729.Sprime LD Drosophila melanogaster embryo POT2 Drosophila
            melanogaster cDNA clone LD47729 Spime similar to X75498;
            D.melanogaster gene encoding proteinA, mRNA sequence.
ACCESSION A1546214 GI:4463587
VERSION   A1546214.1
KEYWORDS EST.
SOURCE    fruit fly.
ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 407)
            Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
            , P., Lewis, S. and Rubin, G. M.
            BDGP/HHMI Drosophila EST Project
            Unpublished (1997)
TITLE     Contact: Harvey, D.
            G. M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 LSA, Berkeley, CA 94720-3200, USA
            Fax: 510 643 9947
            Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
            Plate: 477 row: C column: 5
            High quality sequence stop: 333.
FEATURES
            source
            1..407
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="LD47729"
            /clone_lib="LD Drosophila melanogaster embryo POT2"
            /sex="male and female"
            /dev_stage="0 to 24 hours mixed stage embryonic"
            /lab_host="XLI Blue"
            /note="Organ: embryo; Vector: POT2; Site_1: EcoRI; Site_2:
            XhoI; Sized fractionated cDNAs were directly ligated into
            POT2."
```


KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Mammalia: Eutheria: Primates: Catarrhini, Hominoidea: Homo.
AUTHORS 1 (bases 1 to 460)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ruda Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=MR0&t2=MR0-GN0131-021100-103-b06&t3=2000-11-02&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 460.
Location/Qualifiers
1..460
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GN0131"
/dev_stage="Adult"
/note="Organ: placenta.normal; Vector: puc18; Site_1: SmaI ; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 118 a 89 c 141 g 112 t
ORIGIN
Query Match 72.8%; Score 18.2; DB 172; Length 460;
Best Local Similarity 87.0%; Pred. No. 9,9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 gcacgagtgatcgccagcatt 25
1 ||||| ||||| ||||| |||||
Db 317 GTACGAGGAGCATCGCTGCATT 339
RESULT 43
BE268250 470 bp mRNA EST 13-JUL-2000
LOCUS 601125693F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:345503 5',
DEFINITION mRNA sequence.
ACCESSION BE268250
VERSION BE268250.1 GI:9141852
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
REFERENCE 1 (bases 1 to 470)
NIH-MGC http://mgc.ncl.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgs@bbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov
Plate: LNCM133 row: i column: 24
High quality sequence stop: 470.
Location/Qualifiers
1..470
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:345503"
/clone_lib="NIH_MGC_8"
/tissue_type="Burkitt lymphoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 119 a 129 c 121 g 101 t
ORIGIN
Query Match 72.8%; Score 18.2; DB 165; Length 470;
Best Local Similarity 87.0%; Pred. No. 9,9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 gcgcagagtgatcgccagca 23
1 ||||| ||||| ||||| |||||
Db 227 GCCGACGACGATGCGCAGCA 249
RESULT 44
A1542824 489 bp mRNA EST 22-MAR-1999
LOCUS SD09292.5prine SD Drosophila melanogaster Schneider L2 cell culture
DEFINITION POT2 Drosophila melanogaster cDNA clone SD09292 5prime similar to X75498: D.melanogaster gene encoding proteinA, mRNA sequence.
ACCESSION A1542824
VERSION A1542824.1 GI:4460197
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta: pterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha: Ephydroidea: Drosophilidae: Drosophila.
REFERENCE 1 (bases 1 to 489)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein, P., Lewis, S. and Rubin, G.M.
BDGP/HMT Drosophila EST Project
UNPUBLISHED (1997)
CONTACT: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
Plate: 92 row: H column: 8
High quality sequence stop: 459.
Location/Qualifiers
1..489
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD09292"
/clone_lib="SD Drosophila melanogaster Schneider L2 cell culture POT2"
FEATURES
source

```

/lab_host="DH5-alpha"
/notes="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized
fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
BASE COUNT      109 a      164 c      129 g      87 t
ORIGIN

```

```

Query Match      72.8%; Score 18.2; DB 21; Length 489;
Best Local Similarity 87.0%; Pred. No. 1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY      1 ggcgcagcagtgcatgcgcagca 23
          ||||| ||||| ||||| |||||
Db      427 GCTCAGCGGGGTGCAGCGCCAGCA 449

```

RESULT 45

```

BE018236      504 bp      mRNA      EST      06-JUN-2000
LOCUS      db77d06.y1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3048395.5'
DEFINITION      similar to gb:X57435_cds1 TRANSCRIPTION FACTOR AP-4 (HUMAN);, mRNA
sequence.

```

```

ACCESSION      BE018236
VERSION      BE018236
KEYWORDS      EST.

```

```

SOURCE      human.
ORGANISM      Homo sapiens

```

```

REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE      1 (bases 1 to 504)
JOURNAL      NIH-MGC http://mhc.nci.nih.gov/.

```

```

COMMENT      National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC

```

```

CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
image.linl.gov/image/html/lresources.shtml
Seq primer: -40RP from Gibco
High quality sequence stop: 435.

```

```

FEATURES
Source      Location/Qualifiers
1..504

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3048395"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/notes="Organ: cervix; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."

```

```

BASE COUNT      113 a      162 c      159 g      66 t      4 others
ORIGIN

```

```

Query Match      72.8%; Score 18.2; DB 162; Length 504;
Best Local Similarity 87.0%; Pred. No. 1e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

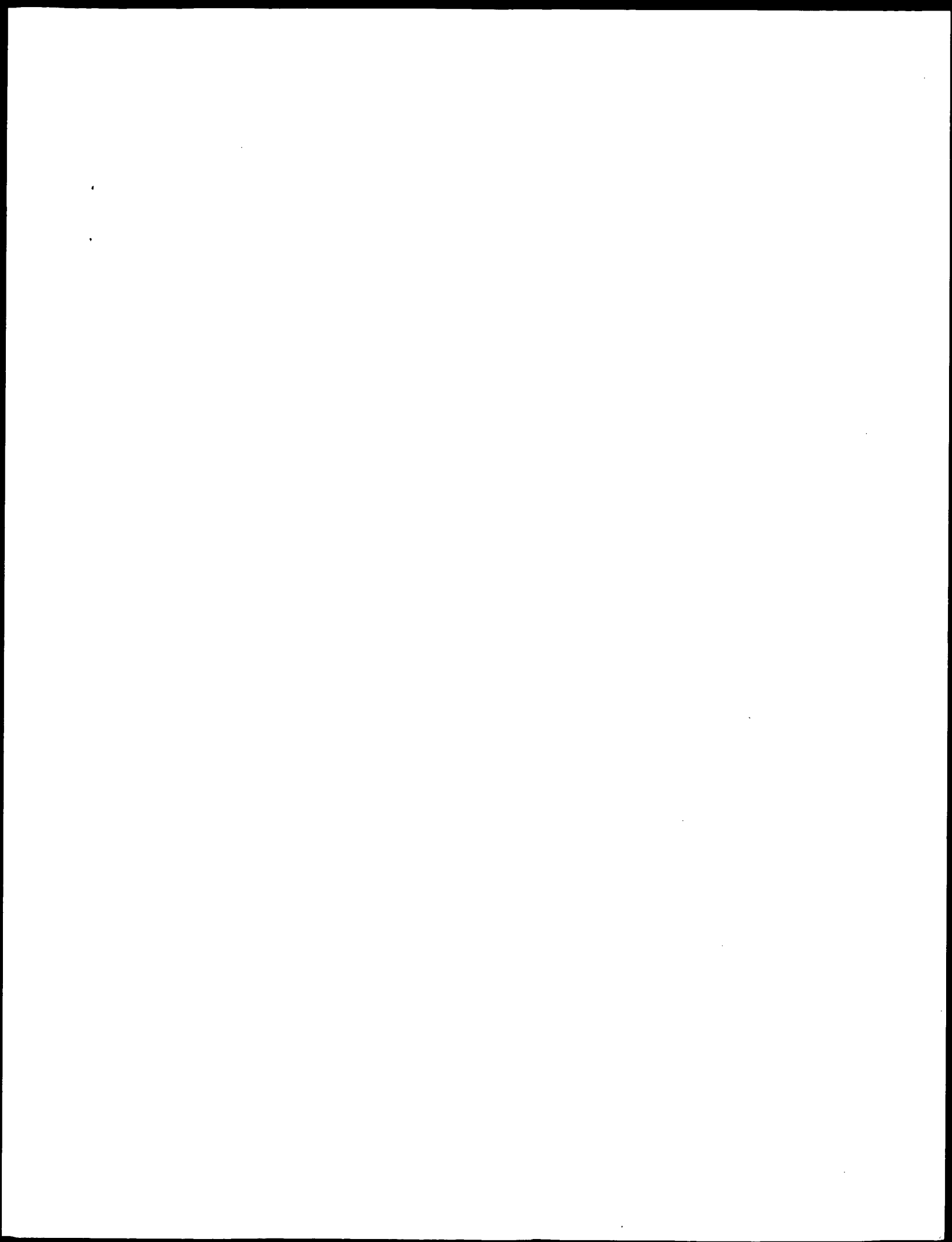
QY      1 ggcgcagcagtgcatgcgcagca 23
          ||||| ||||| ||||| |||||
Db      219 GCGCAGCAGGTGCAGCTGCAGCA 241

```

```

Search completed: October 9, 2001, 13:46:40
Job time: 9515 sec

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 12:11:55 ; Search time 1666.31 Seconds
(without alignments)
232.066 Million cell updates/sec

Title: US-09-396-196f-6

Perfect score: 25
Sequence: 1 gcaggtgcatgccagcatcgcacat 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_ba1:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rtd:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rtd:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_v1:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_v11:*

59: gb_v12:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_pr10:*

95: gb_pr11:*

96: gb_pr12:*

97: gb_pr13:*

98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	9	AR029499	AR029499 Sequence
2	25	100.0	1041	9	AR034916	AR034916 Sequence
3	25	100.0	1084	9	A11530	A11530 B10B gene O
4	25	100.0	1121	10	E00893	E00893 Genomic DNA
5	25	100.0	5793	2	ECOB10	J04423 E.coli 7, 8-
6	25	100.0	5872	9	A38246	A38246 Sequence 1
7	25	100.0	5872	9	A38251	A38251 Sequence 6
8	25	100.0	5872	9	A93674	A93674 Sequence 1

```

9      25      100.0      5872      9      A93679
10     25      100.0      5872      9      AR101809
11     25      100.0      5872      9      AR101810
12     25      100.0      11022      1      AE000180
13     25      100.0      13501      1      AE005258
14     25      100.0      297816      2      AP002553
15     23.4      93.6      5526      2      AF250776
16     20.8      83.2      965      2      AF250770
17     19.4      76.8      10593      1      AE003860
18     19.2      76.8      193351      73      AC068669
19     19      76.0      3603      3      PSEHPRIA
20     18.8      75.2      3148      2      AF192795
21     18.8      75.2      134058      12      AC037425
22     18.6      74.4      11639      1      AE000198
23     18.6      74.4      11641      1      AE005286
24     18.6      74.4      12438      1      AE004044
25     18.6      74.4      13832      2      D90734
26     18.6      74.4      200631      70      AC026813
27     18.6      74.4      212908      73      AC068663
28     18.6      74.4      327773      2      AP002554
29     18.6      74.4      347750      2      AP002998
30     18.4      73.6      119797      63      AC013236
31     18.4      73.6      264853      5      AE003644
32     18.2      72.8      300962      6      DNOSADH04
33     18.2      72.8      9239      2      DB5415
34     18.2      72.8      10864      1      AE004631
35     18.2      72.8      31859      3      MTCY22D7
36     18.2      72.8      103920      63      AC014955
37     18.2      72.8      281869      4      AE003544
38     18      72.0      2923      3      MAV250020
39     17.8      71.2      479      54      G01446
40     17.8      71.2      3748      3      MS031280
41     17.8      71.2      3748      56      CV039574
42     17.8      71.2      35928      6      CELF41H10
43     17.8      71.2      53067      65      AC019940
44     17.8      71.2      68419      64      AC016014
45     17.8      71.2      110000      80      AL359456_2

```

ALIGNMENTS

```

RESULT 1
LOCUS      AR029499      1041 bp      DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION  AR029499
VERSION     AR029499.1      GI:5941472
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 1041)
AUTHORS    Patton,D.Andrew.
TITLE      Enhanced biotin biosynthesis in plant tissue
JOURNAL    Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
SOURCE      1. 1041
            /organism="unknown"
BASE COUNT  262 a      273 c      305 g      201 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1      gcagggtcgcgcgcagcattcgat 25
          |||
Db      84      GCAGGTGCATCGCAGCATTTGCAT 108

```

RESULT 2

```

AR034916
LOCUS      AR034916      1041 bp      DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION  AR034916
VERSION     AR034916.1      GI:5950521
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.
REFERENCE   1 (bases 1 to 1041)
AUTHORS    Patton,D.A.
TITLE      Transgenic plants having increased biotin content
JOURNAL    Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
SOURCE      1. 1041
            /organism="unknown"
BASE COUNT  262 a      273 c      305 g      201 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1      gcagggtcgcgcgcagcattcgat 25
          |||
Db      84      GCAGGTGCATCGCAGCATTTGCAT 108

```

```

RESULT 3
LOCUS      A11530      1084 bp      DNA
DEFINITION Bior gene of E.coli with primers.
ACCESSION  A11530
VERSION     A11530.1      GI:490218
KEYWORDS    Escherichia coli.
SOURCE      Escherichia coli.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia coli.
REFERENCE   1 (bases 1 to 1084)
AUTHORS
JOURNAL    Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES
SOURCE      1. 1084
            Location/Qualifiers
            gene
            CDS
            /gene="Bior"
            /gene="Bior"
            /codon_start=1
            /transl_table=1
            /protein_id="CA00965.1"
            /db_xref="GI:490219"
            /translation="MAHRPMTLSQVTELEFEKPLDULFEAOVHROHPROYAST
            LLSIKTGACPEDCKKYCPSSRYKGLAEIRIMEVOYLESARKAKAGSRPMGAM
            KNPERDMPYLEOMVGVKAMGLACMTLGLTSSQORLANAGLDIYNNIDTSPF
            VPIMLVKVGTPLADNDVDAFDFITFIIVARIIMPTSVRLSAGREONNEOTQAMC
            FMAGANSIFYGCKLLTTPNPEEDKDLQLFKRLGIMPQATVVLADGNEOQORLQALMT
            PTDDEYVNAAL"
BASE COUNT  271 a      286 c      318 g      209 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1      gcagggtcgcgcgcagcattcgat 25
          |||

```


Db 107 GCAGTGCATCCGACGATTTCGAT 131

RESULT 4
E00893 1121 bp DNA PAT 29-SEP-1997
LOCUS Genomic DNA encoding biotin Synthetase.
ACCESSION E00893.1 GI:2169154
VERSION JP 1986149091-A/1.
KEYWORDS Escherichia coli.
SOURCE Escherichia coli.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 1121)
AUTHORS Hirono, Y., Kojima, T. and Kimura, H.
TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN
JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
NIPPON SODA CO LTD
COMMENT OS Escherichia coli
PN JP 1986149091-A/1
PD 07-JUL-1986
PF 24-DEC-1984 JP 1984272605
PT HIRONO YOSHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N15/00,C12N1/20,C12P13/18,C12N1/20,C12R1:19,C12P13/18, PC
C12R1:19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=Escherichia coli Ns101;
CC Feature is identified by experimental;
FH Key Location/Qualifiers
FT CDS 42..1079
FT /product="biotin synthetase".
FEATURES
Source location/Qualifiers
1..1121
/organism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT 289 a 296 c 325 g 211 t
ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 1121;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gcaggtgcatccgacgatttcgat 25
|||||
Db 125 GCAGTGCATCCGACGATTTCGAT 149

RESULT 5
ECOBIO 5793 bp DNA BCT 28-FEB-1994
LOCUS E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase
DEFINITION (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc
protein, and dehydrobiotin synthetase (bioD), complete cds.
ACCESSION J04423.1 GI:145422
VERSION J04423.1
KEYWORDS 7,8-diamino-pelargonic acid aminotransferase;
7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
bioc gene; bioD gene; bioF gene; biotin synthetase; dehydrobiotin
synthetase.
SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 5793)
AUTHORS Otsuka, A.J., Buonocristiani, M.R., Howard, P.K., Flamm, J. and
Johnson, O.

TITLE The Escherichia coli biotin biosynthetic enzyme sequences
JOURNAL U. Biol. Chem. 263, 19577-19585 (1988)
MEDLINE 89066784
COMMENT Draft entry and computer-readable sequence [1] kindly submitted by
A. Otsuka, 09-NOV-1988.
FEATURES
Source location/Qualifiers
1..5793
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
complement(98..574)
/note="ORF 1"
/codon_start=1
/transl_table=1
/protein_id="AA23513.1"
/db_xref="GI:455168"
/translation="MKLSINDLRDGLKLRHVFNGKYGCDNISPLANDVPAKTK
SFVATCYDPAATGSGMHWVYVNLPAIDRYLPDGFSGGLVAMPDGLVLTDTGKTKG
YDGAAPPKGETHRYIFTYHALDIERIDYDEGASAAVGFVHFHSLASATFAMS"
complement(633..1925)
/gene="bioA"
complement(633..1925)
/gene="bioA"
/codon_start=1
/transl_table=1
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AA23514.1"
/db_xref="GI:457106"
/translation="MTTDLAFLDGRHIMHPYTSMTSPPLPVVSAEGCELLISDGR
LVDSGSSWMAAIGHYVNHPLNAAKSSQIDAMSHVFGGITHAPALELCKLVAMSGN
ALECVFSLSSGSAVEVEMKMALOYWAKGASRQRLTFRNGYGDTEFGALMVCDPN
SMHSIMKGYLPENLEAPAPDSRMDEMERGVGFARLMAHREHIAVIEEPYVGA
GCMRHYHEMLAKRIKTRICDRGILLIDELTAGRGGRGRLFACEHAFLADILCKGA
LTGGTMTLSATLTREVAETTSNKGACFMHGPPFGNPLACAAANSALALBEGMQ
OOVADIEVQLREOLAPARDAEMVADVRLVIGAVETTHPVNMAALCKFVEGQWIR
PFGKLIYLPPIYILPQILRLTAVNRAVODETFEFC"
2012..3052
/gene="bioB"
2012..3052
/gene="bioB"
/codon_start=1
/transl_table=1
/product="biotin synthetase"
/protein_id="AA23515.1"
/db_xref="GI:145425"
/translation="MAHRPMTLSQVTELEKPLDLLFEAQYHROHDFPROVST
LISIKTAGCPEDCKYCPQTSRYKTGLEERLEMEVQVLESARKAAAGSTFCGAAM
KNPHERDMPYLEOMVGVKAMGLEACMTLGLTSEAOQLANAGLDYNNHMLDPSPE
YGNITTRVYQERLDTELEKVRDAGIKVSGGLVIGETVKKRAGLLQLANLPTPES
VPIIMLVKVGKGPLADNDVDVAFDRIITIAVATIMPSTSYRLSAGROMEOQAAC
PMAGANSIFGCKLITTPNEPDRDLQLEFKLGLNPQGTAVLADNEDQQLBQALMT
PDTDEYVNAAL"
3049..4203
/gene="bioF"
3049..4203
/gene="bioF"
/codon_start=1
/transl_table=1
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AA23516.1"
/db_xref="GI:145426"
/translation="MSWOKIINALDARRADALRRYVPAVAGAGRMVLVADROLYNF
SSNDYLGSHHQIIRANOCGAEORGISGSGHSGSVYHQAEELEALMELGSR
LITISGFANQVAVIAMMAKEDRIADRLSHASLSLEASLSPSQALRRANDYHLAR
LLAPCGQGVNVTGEGVSMDGDSAPLAEIQDVTOYHNGHMLVDDAHGTGYIGBGRG
SCMLQVXKPELVITFGKGFVGAACVSTVADYLIQFAHLLIYSTMPAQAL
RASLAVIRDEGDARREKIALILTRFAGVGLDPTLADSCSALDPLIYVGNRSALQ
AKRLQGGCWVAIRPVPVAGTARLRLTFAAHMQIDRLLEVLHNG"
4190..4945
/gene="bioc"
4190..4945
/gene="bioc"
/gene="bioc"

```

/codon_start=1
/transl_table=11
/protein_id="AA23517.1"
/db_xref="GI:145427"
/translation="MATVKNQAIIAAFGRAAHYBOHADLOROSADALLAMLPORKYT
HYLDAGCGGMSRHRERHRAQVATLDSLPPALVOAROKDADADHYLADITSLPATA
TFDLAMSNLAVOMCGNLSTALRELYRVYRPKGVVAFPTLLYOGSILPERIQANQVDERP
HANRFLPDEILOSINGVYHOHIOTPIILMPDDALSARSLKGTGATLHGRDRILL
TRSOLOLQIOLAMPQOOGKRPITYHFLGVIARE"
4938..5597
/gene="biob"
/locus_tag="4938..5597"
/gene="biob"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AA23518.1"
/db_xref="GI:145428"
/translation="MSKRYFVTGDTTEVKTIVASCALLQAAKRAGYRTAGYKPYVSGS
EKTPEGLRNSDALORNSIOLDYATVPYTPAEPYSPHIIISOEGRPTEIYMSNG
LRALRQADWVLYEGAGCMFTPLSPTFEADWQVQOLPVLIVGVKLGCTINHAMLIA
OYIOHAGLTLQWVANDYTPPOKRAEYMTTLTRUIPRCWEKSPGLQIKMKOPES
"
BASE COUNT      1363 a 1554 c 1631 g 1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match      100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gcaggtgcacgcacgacatcgat 25
Db      2095 GCAGGTGCATCGCAGCATTCGAT 2119
|||||
|||||

RESULT 6
LOCUS      A38246      5872 bp      DNA
DEFINITION      Sequence 1 from Patent WO9408023.
ACCESSION      A38246
VERSION      A38246.1 GI:2294844
KEYWORDS
SOURCE
ORGANISM      Escherichia coli.
                Escherichia coli
                Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
                Escherichia.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL      Patent: WO 9408023-A 1 14-APR-1994;
                LONZA AG (CH)
COMMENT      Other publication PL 308301 950724
                Other publication CA 2145400 940414
                Other publication AU 4820293 940426
                Other publication HU 71781 960228
                Other publication SK 42095 951108
                Other publication CZ 950809 950913
                Other publication FI 951547 950331
                Other publication JP 85016947 960227.
                Location/Qualifiers
                1..5872
                /organism="Escherichia coli"
                /strain="DSM498"
                /db_xref="taxon:562"
                /clone="PBO30A-15/9"
                1..96
                /function="PROMOTER PTAC"
                /evidence=experimental
                23..28
                /standard_name="PROMOTER PTAC"
                /evidence=experimental

FEATURES
source
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
/standard_name="PROMOTER PTAC"
/evidence=experimental

RBS
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/number=1
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAA02325.1"
/db_xref="GI:2294845"
/translation="MAHRPMTLSQVTELEFERLDDLPEAQOVHROHFDPROVOYST
ILSIKTAGCEDCKYCPOSSRYKTELEFRIMEVEQVLESARAKAAGSTRRCMAAV
KNPHERDMPYLEQMOGVKAMGLEACMTIGTISESQAOLANAGLDYIHNIDTSPER
YGNITTRYQERLDITLEKVRDAGITVCSGAVIGETVKDRAGLLQLANLPTPEES
VPIIMLVKRGTEPLADNDVDARDFTRTIRKAIIMPSTVYRLSAGREQMBOQYAMC
EMAGASITVGCGLTTPNDEEDKDLQTRKIGLNPQOAVIAGDNEQOORLEQALMT
PDDEYVNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATVKNQAIIAAFGRAAHYBOHADLOROSADALLAMLPORKYT
HYLDAGCGGMSRHRERHRAQVATLDSLPPALVOAROKDADADHYLADITSLPATA
TFDLAMSNLAVOMCGNLSTALRELYRVYRPKGVVAFPTLLYOGSILPERIQANQVDERP
HANRFLPDEILOSINGVYHOHIOTPIILMPDDALSARSLKGTGATLHGRDRILL
TRSOLOLQIOLAMPQOOGKRPITYHFLGVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTDILAFDORHIMHPTSMTPUPVYVVSAGCELIJLSDGR
LVDGSSWMAALHGYNHPOLNAAKSOIDAMSHVFGITPAIETLCRTKYANFPPL
LECVLADSGSAVAVVAMKMALOYWAQAGEAROFILFRNGYHGDITGANSYCPDMS
MNSLTKGILPENLFRAPQOSMDGEMDRDVGFRALMAHRHEITAAVILEPVGAGS
GMRKTHPEWLKRIKICDREGILLIADIEATGFGRTGKLFACEHAETIAPDILCGKAL
TGGVTTLSATLTTRVETIISNGAGCFMAGPTFMGNPLACAAANASIALILIESGDMQO
QVADIEVOLREOLA PARDAENVADVVRVIGAGIVETTHPYVMALQCFEYQGVWIRP
FGKLIYMPPIYIILPQOLQRLTAAVNRAVQDETFFCQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6

```

```

/evidence-experimental
/product="PROTEIN"
/db_xref="GI:2294848"
/translation="MKLISNDLRDGLPHRRHVTNGMGYDGNISPHLMDVPAKTK
SEVYCYDPAPTGSGMMHVVVNLPAATRYLPQSGSLVAMPDGYLOTPTDGKTC
YDGAAPRKGEHRYIFTYHALDIERIDVDGASGMVGFNVHPSLASASTAMES"
stem_loop
5583..5605
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 gcaggtgcacgcgcacattcgat 25
        |||
Db      200 GCAGGTGCATCGCCAGCATTTGCAT 224

RESULT 7
LOCUS      A38251      5872 bp      DNA
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION  A38251
VERSION     A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
            Escherichia coli.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL    Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960328
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
            location/Qualifiers
FEATURES
source
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..22308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CA02329.1"
/db_xref="GI:2294850"
/translation="MSWOKINNALDARRAARADLRRTYVAGACGMLVADDDRYNF
SSNDYGLSHHPDITRAWOGAEOFGISGGSGHVSYSVHDALEELAEWLGYSRA
LLEFISGFANQAVIAAMAKEDRIADRLSHASILEASLSPSOLRRFAHNDVTHLAR
LIASPCGOOMVTEGVFSMDGSDAPLAEIOGVTOOHNGMLWMDAHDGTVIGEOGRG

```

```

SCWLOKYPPELLVVTGKGFGVSGAALVCSSTVADYLLQFARHLIYTSMPAQOAL
RASLAVRSDEGDARREKLALITRRAGVODLPFTLADSCAIOPLIIVGDNRSALOL
AEKLRQGCWVTAIRPPTVAGTARLRILTLTAHEMQDDIRLLEVHNGS"
RBS
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/standard_name="DEHYDROBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYFVTGDTDEVEGKTVASCALLQAAKAGYRTAGYKPVASG
EKPEGLRNSDALALORNSSIOLDYATVNPYTAEPSPHIIISAQGRPTESIVMSAG
LRALFOQADWVIVGAGCGMPPLSDPFTFADWVTOROLPVILVGVKIGIINAMLLTA
OVIQHAELITLAGVANDYTPPKRHAETMTTLRMIPAPLGRIPMLAENPEAAATGK
YINLAEVDASTLGEFTSRL"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 gcaggtgcacgcgcacattcgat 25
        |||
Db      200 GCAGGTGCATCGCCAGCATTTGCAT 224

RESULT 8
LOCUS      A93674      5872 bp      DNA
DEFINITION Sequence 1 from Patent EP0798384.
ACCESSION  A93674
VERSION     A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
            Escherichia coli.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch, O. and Brass, J.
TITLE      Biotechnological method of producing biotin
JOURNAL    Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
COMMENT     location/Qualifiers
FEATURES
source
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
/evidence=experimental
45..49
/standard_name="PROMOTER PTAC"
/evidence=experimental
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1

```

```

/transl_table=11
/evidence-experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPWTLSOVTEHPKPLLDLFEAOVYHROHFDPROVOST"
LSTIKGACPEDECKYCPOSSRYKTGLABERLMEVEQVLESARKAKAGSTRFCWGAAM
KNPHRBDPTEQWQVKGAMGLACVATLGLTSESQAQRLANAGLDYNNHLDSPFEF
YGNITTRTYOERLDTLEKVDAGIKVCSGIVGLGTVDRAGILLOLANLPPTPES
VPIMLVYKGTPLADNDVDADFEIPLTINAVARIMPTSVYRLSAGREMEOTQAMC
FMAGANSIFYCGKLLTTPNEDEKDIQLFRRLGTLNPQOTAVLAGDNDNQQRRLQALMT
PDTEYVNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVNRKQALAAFGRAAHYEQHADLROSADALLAMPQRYT
HYLDAGGPGMSHMRERHAAOYATLDSPMIYQAKRQADHYLAGDIESLPLATA
TFDLAMSNLAVMQNGNLTALRELYRVKRGVAFITLVQSLPELHQWQVDERP
HANFLEPDELEOSLVGHYOHIIQPTTFMDALSMNSLKGIGATHLHEGRDPRIL
TRSOLOQLQALAMPQOGRPLTYHLFLGVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANONATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAFDPORHIMHPYTSPLPYVPVYSAEGELLISGR
LVGSSWMAAHTGYNNHNPOLNARSIDAMSVYMGCTTHAPTELCRLVAMPPOP
LECVPLADSSAVVEYAMKALLOYKAKGEARORELTFRNGHGTGKMSVCDPDS
MSLWKGYLPENLFAPAPOSRMDEWDEDMGFEARLMAHREHIAVITIEPIVQAG
GMKMPHEWLKRLIKICDREGILLADEIATGFGTGLFACENHAIADIDILGKAL
TGGTMTLSATLITREVAET ISNGEACGFMHGPFGNPLCAANAASIALILSGDMQO
QVADIVOLREOLAPARDAEMVADVRLGAGVETTHVYVNAALOKFVEGQWIRP
FGKLILMPYITLLPQQLRLTAANVRAVODEFFQO"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRQDCKLPHRHVFNMGVGDGNISPHLAMDVPAGTK
SRVVTCTDPDAPTGSGMHWVNLPAIDRVLPQSGGLVAMPDGVLOTTRDCKTG
YGAAPKGTETRYLFTVHALDIERIDVDEGASGAMGVNFHSLASAITAMS"
5583..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
5583..5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
5552 C 1695 G 1307 T

```

```

ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcaagtcacgcacgattcgat 25
DB 200 GCAAGTCACATGCCAGCATTCGAT 224
RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
source location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQKXTNALDARADALRRPYVAGAGRLVADKROYLNF
SSNYLIGSHHPQILIRAMQOAGEQFGSGGSHVSGYSVHQLEBELAEWLGYSNA
LFTSGFANQAVIAAMAKEDRIADRLSHASILEASISPQLRFRANDVTHLAR
LHSPCRQOQWVTEGVFSMDGSDAPLAEIOQVQHNGLMWDAGTGVIEGCGG
SCWLQKVPBELLVVTFGKFGFSGAVALCSYADVLLQFARHLIYSNPPAOAOL
RASLAVIRSDGDAERKLAALITRFPAVGOQDLPETLADSCSAIOPITVGDNRALOL
AEKLROGQWVTAIRPPTVPAGTARLRLTLTAHEMDIDRLLEVLHGNG"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYFVGTDTDEVGKTVASCALLQAAKAGRYTAGYKPVASG
EKTPEGLRNSDALQNRNSSLQDVTAVTVNYPYFAETSPHIIISQGRPIESLVMASG
LRALQDQWVLEVGAGWETPLSDTFTEADWVTOQLPVLIVGVKLCINHAMLTA
QVIOHAGLTLGAVVANDVTPPGKRHAIEYMTTLTRMIPAPLIGETPMLANPEMAATGK

```

BASE COUNT 1318 A 1552 C 1695 G 1307 T

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 YINLAFVDASTIGFTSRU"

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.38;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcaggtgcatgcgcagcatcgcattgcg 25
 ||||||||||||||||||||||||||||
 DB 200 GCAGGTGCATGCCAGCATTTGCAT 224

RESULT 10
 LOCUS ARI01809 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 1 from patent US 6083712.
 ACCESSION ARI01809
 VERSION ARI01809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 FEATURES
 source 1. 5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.38;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcaggtgcatgcgcagcatcgcattgcg 25
 ||||||||||||||||||||||||||||
 DB 200 GCAGGTGCATGCCAGCATTTGCAT 224

RESULT 11
 LOCUS ARI01810 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION ARI01810
 VERSION ARI01810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
 FEATURES
 source 1. 5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.38;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcaggtgcatgcgcagcatcgcattgcg 25
 ||||||||||||||||||||||||||||
 DB 200 GCAGGTGCATGCCAGCATTTGCAT 224

RESULT 12
 LOCUS AE000180 11022 bp DNA BCT 01-DEC-2000
 DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.
 ACCESSION AE000180 U00096
 VERSION AE000180.1 GI:1786988
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Escherichia coli K12.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J.J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503

REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett, G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: mark@ember.gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES
 source Location/Qualifiers
 1..11022 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"


```

VERSION      AE005258.1 GI:12513751
KEYWORDS
SOURCE       Escherichia coli O157:H7 EDL933.
ORGANISM     Escherichia coli O157:H7 EDL933
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE    1 (bases 1 to 13501)
AUTHORS      Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Postell, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Welch, R.A. and Blattner, F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
TITLE        Nature 409 (6819), 529-533 (2001)
JOURNAL      MEDLINE
PUBMED      21074935
REFERENCE    2 (bases 1 to 13501)
AUTHORS      Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Postell, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Welch, R.A. and Blattner, F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source       1. 13501
Location/Qualifiers
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
<1. 7576
/note="O-island #36; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655; Cryptic prophage
CP-933K; Includes one copy of the 13 bp direct repeat that
flanks the prophage"
66. 665
/gene="lomK"
/note="20981"
66. 665
/gene="lomK"
/function="lomK"
/note="putative membrane; Other or unknown (phage or
prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to
residues 1 to 199 of 199 from Genpept 118 :
g1175327891gb|AA63231.1|AF151091_2 (AF151091) lom
[prophage P-Elba]"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein of prophage
CP-933K"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MKRVCAAILSAICLAVSPVPMASEHOSTLSAGYLHASTAPG
SDILNGINVKRYREFTDITGLIFSYANAEDOKTHYSPTRMHEDYVRNRFVSVAG
PSYRVNEMFSAVMAGVAYSRVTSFGSDYFRYDNKRKTHDYLITGSDDARYSNTSLAW
GAGVQFNPTESSVAVDAVAYEXSGSDMRTDQFIVGVYKF"
719. 2041
/gene="20982"
719. 2041
/gene="20982"
/function="20982"
/note="putative structure; Structural component (phage
or prophage Related)"
/note="Residues 164 to 440 of 440 are 68.79 pct identical
to residues 381 to 645 of 645 from Genpept 118 :
g1145854361gb|AAD25464.1|AF125520_59 (AF125520) putative
tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="putative tail component of prophage CP-933K"

/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAVOISGVLKDGAKPKIONCTLOKARRNSTYVYVTVASE
NPDEAGRYSMDEVYGOYSVTLVGEFPPSHAGTISYEDSQPTLNDPFGAMTEDAR
PEARLRFEOMVEEARHAEAKKNAAGAEISARNAQISAKASASAMATEDADSE
SARQASASASAKSESEASSASSAEAKKSESIQSQTDAELSKTAESAGNAADA
TTSREKAREASAEOSRISAEADVNNITPTVYGPPGPGGPGGPGGPGGPGG
RGDTGPGATGTERGPGGDTGPGAGCGKCGRGERGEMGLGNMGPGGPGGTCGA
PGGPGKSTGAGVGVATGPGGPGGDEETIIFRLGPKMTITENSGMPGPGDGLI
TGLTFLEPKATQVQGMFQHLQVRFSGDPMDVKGDLDEVSDFGRTE"
2004. 2312
/gene="20984"
2004. 2312
/gene="20984"
/function="orf; Other or unknown (phage or Prophage
Related)"
/note="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 of 129 from Genpept 118 :
g1145854371gb|AAD25465.1|AF125520_60 (AF125520)
hypothetical protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MKNAVIOALELNDMNILRLKMLQSLGGCKHDCENGSRSLTAOLR
LGPADLSEDENGILPEODRVTIYVILADKQKQICVVRPLQILRADGTWENIGMK
"
2489. 3469
/gene="20985"
2489. 3469
/gene="20985"
/function="orf; Other or unknown (phage or Prophage
Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from Genpept 118 :
g116903671gb|AA63232.1 (AF101016) hypothetical protein
predicted by glimmer [Salmonella typhimurium Lr2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPITRPFNHSVNIQVSPQOVSFAGKEYLKVIDEKTPIIF
OMFEPNPERYKDEYPIVNTKQHPYLNTAARLSESPRMGITGVDFGDSNOKTAS
KLERDFENMIIYREDVDFSMTRKLSITIDITCEQRLTDEKRDYLLMLLKELR
EISKQDSLISMYAKRRHAWDFERNLALIKAGIEFCTYNTNKHGISFEGCTYLD
MMILITGLKGTIYAPDGISMVDRNDSDVNTENSAILVNRSHPALLEGSLSPMSKVD
AHYVDGLGKGVKKYFNFTPLHNYNHCFDFTIEFNHPIIMNTSQTCSW"
3530. 4522
/gene="20986"
3530. 4522
/gene="20986"
/function="orf; Other or unknown (phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPISLQSNFNSAPAGYAPFIAPRANAAYADYVLDIGRKRIPL
SAADSLNYEESYIRAVHDSRSLIDOHVDMGIVDALSRSGFRVAVSGIINEK
VHIGCKIRNETELNEBESSVKIDIDQSLTCELYDYDGOEPIFPCAGRNDPEPY
VFSVAPDSDSEMPSSWQGLTHEIITHVTSDDSSGDSNIBLGFETELRLVAQDEL
EYSASVNIQYGFIOQHDFPGAINDLQDANOIOLYHGAPYITFTGVDKHNQO"
5350. 6231
/gene="20989"
5350. 6231
/gene="20989"
/function="orf; Other or unknown (phage or Prophage
Related)"

```

```

/Note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG5142.1"
/db_xref="GI:12513757"
/translation="MISPSYVIGCGSNLSLRNLITSPDNRLVSSVRQAAVHSDNGAIV
KGNRTYVATDKNKCYTRRSHSGCFTNLRLGMPKGLISRKIEVLNAPSASAM
ERGLVHNRPLPPVDYAPPELPSVAYNRSNGVNGKAVVADADATKVLAM
FTTSQNEEVTSVCEVCEVQYAGASAEKTYGNNDDIGIRMDKINGESLNLSPAO
AEHAIDMEFDLEQKGLIFVDTETNVLVDRAKNEFNPIDISSYVNSDRSMSEQIMQ
SYHGKODLISVLSKI"
6462..7160
/gene="20990"
/gene="20990"
/feature="orf: Other or unknown (Phage or Prophage
related)"
/Note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG5143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNVLHSSRSSMSDPTDIESLVASSVQWTKNNPOLRF
OGTDNITVOGIEALDKIGSETGRVLLNIESISRLKSTVLIHLNLSPLCYMAHD
IDANHRGTSGDFPCNNAVYPCGEGISYVDPAITVFHELLHAFHNIGERIKVES
SRAOSKYSPLLEARTVGLGAFSEVLSSENKFHEITGMPRTSTPDSALIHDDNT
VSLGQVRLHPL"
complement(7668..8144)
/gene="ybhb"
/Note="20992"
complement(7668..8144)
/gene="ybhb"
/function="orf: Unknown function"
/Note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG5144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGLKLRHRAVNGMGYDGDNIISPLAMDVPACTK
SYVATCPDPAPGSGMMWVNLPAIDRVLPQGGSGIVAMPDGLVTRDFEGKIG
YDGAAPKGETHRYIFTVALDIVERIDVDEGASGAMVGENVHPHSLASITAMPS"
complement(8203..9492)
/gene="b10A"
/Note="20993"
complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcaggtcatcgccagcattcgat 25
Db 9662 GCAGGTGATCGCCAGCATTTTCAT 9686
RESULT 14
AP002553
LOCUS AP002553 297816 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AF002553 BA000007
VERSION AF002553.1 GI:13360211
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (sites)
/Note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG5142.1"
/db_xref="GI:12513757"
/translation="MISPSYVIGCGSNLSLRNLITSPDNRLVSSVRQAAVHSDNGAIV
KGNRTYVATDKNKCYTRRSHSGCFTNLRLGMPKGLISRKIEVLNAPSASAM
ERGLVHNRPLPPVDYAPPELPSVAYNRSNGVNGKAVVADADATKVLAM
FTTSQNEEVTSVCEVCEVQYAGASAEKTYGNNDDIGIRMDKINGESLNLSPAO
AEHAIDMEFDLEQKGLIFVDTETNVLVDRAKNEFNPIDISSYVNSDRSMSEQIMQ
SYHGKODLISVLSKI"
6462..7160
/gene="20990"
/gene="20990"
/feature="orf: Other or unknown (Phage or Prophage
related)"
/Note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG5143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNVLHSSRSSMSDPTDIESLVASSVQWTKNNPOLRF
OGTDNITVOGIEALDKIGSETGRVLLNIESISRLKSTVLIHLNLSPLCYMAHD
IDANHRGTSGDFPCNNAVYPCGEGISYVDPAITVFHELLHAFHNIGERIKVES
SRAOSKYSPLLEARTVGLGAFSEVLSSENKFHEITGMPRTSTPDSALIHDDNT
VSLGQVRLHPL"
complement(7668..8144)
/gene="ybhb"
/Note="20992"
complement(7668..8144)
/gene="ybhb"
/function="orf: Unknown function"
/Note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG5144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGLKLRHRAVNGMGYDGDNIISPLAMDVPACTK
SYVATCPDPAPGSGMMWVNLPAIDRVLPQGGSGIVAMPDGLVTRDFEGKIG
YDGAAPKGETHRYIFTVALDIVERIDVDEGASGAMVGENVHPHSLASITAMPS"
complement(8203..9492)
/gene="b10A"
/Note="20993"
complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.36;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcaggtcatcgccagcattcgat 25
Db 9662 GCAGGTGATCGCCAGCATTTTCAT 9686

```

AUTHORS

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
REFERENCE
AUTHORS


```

gene      1245..2114
          /gene="ECS0754"
CDS       1245..2114
          /note="similar to SUCD_ECOLI g111786949 percent identity
          100 in 289 aa (Conserved in E.coli K-12)"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="succinyl-CoA synthetase alpha subunit"
          /protein_id="BAB34177.1"
          /db_xref="GI:13360213"
          /translation="MSILIDKNTKVICOGFTGSGCFHSEDAIANGTKMGVTPGKG
          GTTHGLPENTREAVATGATGASVIYVPAFCSDILAIADIGIKLITITTEIPF
          ADMLTVKXKIDAGYRMIGPNPCGVITTEGCKIGIQPHIKRPGVIGVSSGCTIYE
          AVKQTDYFGSGTGVIGDIPGNSNTIDILEMEKDPOTEALVIGETGSAFEAA
          LADIEHALKTVLK"
          /db_xref="GI:13360216"
          /db_xref="ECS0755"
          /db_xref="ECS0755"
          /db_xref="ECS0755"
          /note="probable transcriptional regulator, similar to
          transcriptional regulators e.g. glycine cleavage system
          transcription activator (gcv operon activator) -
          Escherichia coli g14170431sp|p32064|GCV_A_ECOLI percent
          identity 31 in 300 aa"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="putative transcription regulator"
          /protein_id="BAB34178.1"
          /db_xref="GI:13360214"
          /translation="MRGKIPKTELVTEEVARHSEYTRAEELATGSAVFRVYSAL
          EEFLLRPLFESHKSKRFLENDGKYLGIVKETLNLERDTNIMTQPVQVTELVAN
          PFSTHMLIPNLHEFTKLHPDIIIVNIHSLANGDPLNREDAVINRENFCAWAEVY
          LPEEELIPVCSGSLAMSDOKSLVAELTELPDLHQSTRITGMEEMFALSGVSPVLN
          NGPRDILSMILAAVRSNCGVALPRLFAIHDLDSDGDMVITPCDVPITRGNRIMTQE
          EKSDSPHLQDFREMLAKSVPOEM"
          /complement(3720..4322)
          /gene="ECS0756"
          /complement(3720..4322)
          /gene="ECS0756"
          /note="probable cob(1)alamin adenosyltransferase, similar
          to cob(1)alamin adenosyltransferases (corrinoid
          adenosyltransferases) e.g. [Escherichia coli]
          g11151481sp|p3040|BVR_ECOLI percent identity 67 in 200
          aa"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="putative cob(1)alamin adenosyltransferase"
          /protein_id="BAB34179.1"
          /db_xref="GI:13360215"
          /translation="MEARSTERRHVOOROKLKEOYDTRVAALTEKKGILIVTNGKG
          KSTARCTYTRAVGKRTYGVAYIKGDMNDENFYNILQVGEFHMIGTGWETQNR
          QADIDAKVWSESKRMADKRDVLVDLELTLYMALHYHLDIEVYIASIQNRPACQSV
          IYTRGCHSQILKMDPTSEIRPVKHAIFDNGIQADPGIDW"
          /complement(4332..5984)
          /gene="ECS0757"
          /complement(4332..5984)
          /gene="ECS0757"
          /note="probable fumarate hydratase, similar to fumarate
          hydratases e.g. fumarate hydratase class I, aerobic
          (fumarase) - Escherichia coli
          g11205981sp|p00923|FUMA_ECOLI percent identity 68 in 545
          aa"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="putative fumarate hydratase"
          /protein_id="BAB34180.1"
          /db_xref="GI:13360216"

```

```

          /translation="MSKPIWQEPFLQNKDTEYTLISDQITVELDEEVIKIAPE
          ALTLARQAFYEASFLRSARHLOQVAILINDQASNDKYVALQOLLRNAEVSAKGLP
          NCOQDTTATIVASKGQOIMTGNDAEALISGIYSTROENNLRFSONAPIDMYEVNTO
          TNLPAQDIDISAVADEYHFLCYVKKGGSSAKKALYQETSILQPELTAFLIEKMSL
          GTACQPPYHIAFVVGGLSADQTLKIALASTKYYDNLPTISNGEOGAFDIDELKYL
          EASQOFGIAGFGGKYFAHDIRVIRLRPHGSGCPIDAMALSADRNIRKATKHGILWL
          EKLEHNPGOYIPASLRENNHQAHOYQDLNPLRDVMDLRLPVGRVLSGPIVAVR
          DIAKIKARIDSGEPMPEYKLNHIVYVAGPAPTPERMAGCSLGPPTGGMDGYIDTF
          QAAGSLVMTSKGNRSQOYDACHKHGCFNLISIGGAALIAOEYKSLNCLYPELG
          MEAVWMEVENLPAFLIYVDKGNPFQSGFDOHRCAKSPAGH"
          /complement(6092..7372)
          /gene="ECS0758"
          /complement(6092..7372)
          /gene="ECS0758"
          /note="probable transport protein, similar to
          glutamate/aspartate transport proteins (proton glutamate
          symport proteins) e.g. [Bacillus stearothermophilus]
          g11214671sp|p24943|GLT_BACST percent identity 38 in 416
          aa, also similar to C4-dicarboxylate transport"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="putative transport protein"
          /protein_id="BAB34181.1"
          /db_xref="GI:13360217"
          /translation="MRKISLTMTLLALVLMGIIIGVYLNNTASPEAKLYAOEISFT
          TIRLRIMKITAPLVYSTIVYGIAMKMDAKALGIFPSKTLFPLCASLSIALGILTY
          NFEWPGTGINFVANGAETTGVAAPETTLAVFISHAPTSIVDMAHNETLOIYVSI
          FLGSLTAIGKESAIIVHALDSLAAHAKLGLVYMLPAPLTPRALSAALAEGLAVM
          VSAQIFMGEEFTMLLMVLLIGAIYVGGCIRLRRLRALSEPALSALEFTSSAEAP
          GTLEKLEDFGVSPKIASFVLPIGYSMFNLVSGMAYCSATYFIAQACNHIHSIEQITM
          LLILMTSKGAGVBRASMYVIAITNIPNPEGLILLMGVDFLDMGSRATNVMSN
          AMGAAMVSRMGHREHFGECRCALAKPMSNALP"
          /complement(7533..7850)
          /gene="ECS0759"
          /complement(7533..7850)
          /gene="ECS0759"
          /note="unknown"
          /codon_start=1
          /transl_table=11
          /evidence=not_experimental
          /product="hypothetical protein"
          /protein_id="BAB34182.1"

```

```

Query Match      100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1  gcaggtgcacgcgcagcattcgat 25
Db      92394  GCAGGTGCATGCCAGCATTTCGAT 92418

RESULT  15
AF250776
LOCUS   AF250776 5526 bp DNA BCT 31-JAN-2001
DEFINITION
  Uncultured bacterium pCosHE2 hypothetical 17.1 kba protein in
  modc-b10a intergenic region, DAP-aminotransferase B10a (b10a),
  biotin synthase B10b (b10b), KAPA synthetase B10f (b10f), and
  biotin biosynthesis protein B10c (b10c) genes, complete cds; and
  dehydrobiotin synthetase B10d (b10d) gene, partial cds.
ACCESSION
  AF250776
VERSION
  AF250776.1 GI:12620124
KEYWORDS
  SOURCE
  ORGANISM
  .
  uncultured bacterium pCosHE2.
  uncultured bacterium pCosHE2.
  Bacteria: environmental samples.
  1 (bases 1 to 5526)
REFERENCE
  Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streif, W.R.
  Direct cloning from enrichment cultures: a reliable strategy for
  isolation of complete operons and genes from microbial consortia
  Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
JOURNAL
  MEDLINE
  20575196

```

PUBMED 11133432
 2 (bases 1 to 5526)
 Entcheva, P., Liebl, W. and Strell, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
 FEATURES
 source
 1. 5526
 /organism="uncultured bacterium pCoshE2"
 /db_xref="taxon:143797"
 /clone="pCoshE2"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(52. .528)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical 17.1 kDa protein in modc-bioa
 intergenic region"
 /protein_id="AAG60577.1"
 /db_xref="GI:12620125"
 /translation="MKLISNDLSDGDKLPHRHVFNMGVGDGDNISPHLAWDVPAGTK
 SFVVCYDPAATGSGMWVWVNLPAOTRVLPQFGSGLVAMPDGYLQTRTDFGKAG
 YDGAAPKGETHRYFTVHALDIERIDVDEGASAMGVNFHPSIASSTAMES"
 complement(587. .1876)
 /gene="bioa"
 /complement(587. .1876)
 /gene="bioa"
 /note="7,8-diaminopelargonic acid
 synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase Bioa"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDDLAFDORHIMWPTMTSPLVYVYVSAEGCELLISDGR
 LVDGSSWMAAIGHYVNPOLNAANKSQIDAMSHVFGITIHAPAIETCRKIVAMPPOP
 LKCVLADSGVAEVAAMKMAIQWAKGAEKROFLIFRANGHGDITGAMSCVDPDS
 MSLKGYLPEMLFAPAPQSRMDGMDERDVGAFRLAARHETIAVIEPTVQSG
 GMRMYPEMLKIRIKICREGILLIADIEATPGRTKIFACHEAELIAPOTICGKAL
 TGGMTLSATLTTRVAVETISNGEAGCCMHGPTMGNPLACAAANASLITLSDGWO
 OVADIEVLRQOLAPARAENVADVRLVIGVETTHPVMAALOKFEVQGVIRP
 FGKLTLMPTIILPQOLRLTAANRAVODETEFCQ"
 1963. .3003
 /gene="bioa"
 1963. .3003
 /gene="bioa"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase BioB"
 /protein_id="AAG60579.1"
 /db_xref="GI:12620127"
 /translation="MAHRPWTLSQVTELEFEPKPLDLLFEAQOVHROHPDPROVOST
 LLSIKTGACPEDCKYQSSRYKTGLEMERMEVEYVLESARKAKAGSGSRPGMAW
 KNPNERDMPYLEQWVGKIDGLFACMTIGTISLSSQORLANGLDYNNINLIPSP
 YGNIITRTYQERLDTEKRYDAGIKCSGGITVIGETVDRGRLQLLNLPTPPPS
 VPIMLVKVKGTPLANDVDVDFPITITAVARIMPTSVVRLSAGEQNNEDTOAC
 FMAGANSIFGCKLITTPNPEEDKQLFRLGLNPQOTAVLAGDNEQOORLQALMT
 PDTEYVNAAL"
 3000. .4154
 /gene="bioa"
 3000. .4154
 /gene="bioa"
 /note="8-amino-7-oxononanoate synthase"
 /codon_start=1
 /transl_table=11
 /product="KAPA synthetase BioF"
 /protein_id="AAG60580.1"
 /db_xref="GI:12620128"
 /translation="MSWQEKINAALDARRAALRRRYPAVAGAGRWLVADDDROYLNF
 SSNDYIGLSHPOLIRAMQGAEGFGISGSGSHVSGYVHQALEEELAEMLGYSGRA
 LFTSGAANDAVIAAMAKEDRIADRLSHASLEASISPSQLRRPAPHNDVTHLAR
 LASPCGOQLVTEGVFSMDGSAPLAEIQOVTQOHNGMLAMDADAHGTCVIGEGRC

SCMLQKVPPELLVTFEGKGVSGAAVLCSSVTADYLLQFARHLIYSTMPPRQAOL
 RASLAVIRREDDARREKIVSILARFAGVODLPFLBSCSNIQPLVGDNSRALQL
 AEKLRQOGCVTAIRPTVPAGTARLRLITLTAHEMODIDRLLEVLHNG"
 4141. .4896
 /gene="bioc"
 4141. .4896
 /gene="bioc"
 /note="reaction step prior to synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=11
 /product="biotin biosynthesis protein Bioc"
 /protein_id="AAG60581.1"
 /db_xref="GI:12620129"
 /translation="MATVVKQALIAAFAAGRAALYEQHADLOROSADVILAMPKRYT
 HVLADGPGPMTRHMRERHAYVALDLSPPMUYQAROKDADHYLAGDIESPLATA
 TFDLANSNLAWQMGNSLSTALRELYVRSKGVARTPYVGSIPRELRANQAVDER
 HANRFLPDPDIESLNGVYOHHTIOTITLFPDALSAMSLSKIGATHLHSDRPRIL
 TRSOLRLQILAMPQDQGRYPLTYHLFLGVARE"
 4883. .5526
 /gene="bioc"
 4883. .5526
 /gene="bioc"
 /note="DTP synthetase"
 /codon_start=1
 /transl_table=11
 /product="dethiobiotin synthetase BioD"
 /protein_id="AAG60582.1"
 /db_xref="GI:12620130"
 /translation="MLVSKRYEYVGTDFVGVKTVASCALLQAKAGRYTAGKPVAS
 GSEKTEGRLNSDALIORNSSLQIDYATVNVNTPAEPSPIITSAOEGRIEVSMS
 SGRLAEQDQAVIVGAGCWETPSIDTFEPADWYQOBDLPILVYGVKLGCTINAML
 TQAOTIAGSLTAGVANDVTPPGKHAETITTLRMIPALGEIPWLAENPE"
 BASE COUNT 1274 a 1507 c 1567 g 1178 t
 ORIGIN
 Query Match 93.6%; Score 23.4; DB 2; Length 5526;
 Best local Similarity 96.0%; Pred. No. 2.1;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 gcagtgcatgcgcagcatgcgat 25
 Db 2046 GCAGTCCATGCCAGCATTCGAT 2070
 ||||| ||||| ||||| ||||| |||||
 RESULT 16
 AF250770 965 bp DNA BCT 31-JAN-2001
 LOCUS uncultured bacterium pCoshE1 DAPA-aminotransferase (bioa) and
 DEFINITION biotin synthase (biob) genes, partial cds.
 ACCESSION AF250770
 VERSION AF250770
 KEYWORDS
 SOURCE uncultured bacterium pCoshE1.
 ORGANISM uncultured bacterium pCoshE1.
 Bacteria: environmental samples.
 REFERENCE 1 (bases 1 to 965)
 Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Strell, W.R.
 TITLE Direct cloning from enrichment cultures, a reliable strategy for
 JOURNAL isolation of complete operons and genes from microbial consortia
 MEDLINE Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
 PUBMED 11133432
 20575196
 11133432
 2 (bases 1 to 965)
 Entcheva, P., Liebl, W. and Strell, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
 FEATURES
 source
 1. 965
 /organism="uncultured bacterium pCoshE1"
 /db_xref="taxon:143796"
 /clone="pCoshE1"

gene /note="unknown organism, cosmid clone derived from environmental consortium"
complement(1.438)
/gene="bioA"
/complement(1.438)
/gene="bioA"
/note="7,8-diaminopelargonic acid synthetase-aminotransferase; similar to bioA"
/codon_start=1
/transl_table=11
/product="DAPA-aminotransferase"
/protein_id="AAG60563.1"
/db_xref="GI:12620105"
/translation="MTQDDIAFDQOHIMPTSMRPPYVPAVSAHACORRRRLVD
GKSSWMAIHGTHNPRLNAKAKQIEQSHVWFQDTTHQPAVDLCRRXVAMTPDALEC
VFLADSGSAKXVDENGAATAPLPADRPERRXAXFIFXLQ"
525..>965
/gene="bioB"
525..>965
/gene="bioB"
/note="similar to bioB"
/codon_start=1
/transl_table=11
/product="biotin synthase"
/protein_id="AAG60564.1"
/db_xref="GI:12620106"
/translation="MKHHARMTWSQVTELPNKPPLLEMEFAQOVRHQPDPHNOXST
LSTIKGACPEDECKYCPQSAARYKTGLSESLRMEVQVDSARAKNNAOSTRECMCAAM
KKPHIVICPYLEQVQVKGVEKMGLEALYDPPRAERRAXXSAQLLA"
BASE COUNT 230 a 256 c 285 g 180 t 14 others

Query Match 83.2%; Score 20.8; DB 2; Length 965;
Best local Similarity 91.7%; Pred. No. 38;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 caggtgcatgcagcattcgat 25
|||||
Db 609 CAGGTGCTGTCAGCATTGAT 632

RESULT 17
AE003860 10593 bp DNA BCT 06-JUL-2000
LOCUS Xylella fastidiosa, section 6 of 229 of the complete genome.
DEFINITION AE003860 AE003849
VERSION AE003860.1 GI:9104830
KEYWORDS
SOURCE Xylella fastidiosa.
ORGANISM Xylella fastidiosa
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.

REFERENCE 1 (bases 1 to 10593)
AUTHORS Silvestri, M.L., Siqueira, M.J., de Souza, A.A., de Souza, A.P.,
Terenzi, M.F., Truffi, D., Tsai, S.M., Tshako, M.H., Vallada, H., Van
Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Zago, M.A., Zatz, M.,
Weidman, J., and Setubal, J.C.
The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis, Sao Paulo, Brazil
Nature 406 (6792), 151-157 (2000)
20365717

JOURNAL MEDLINE
REFERENCE
AUTHORS

REMARK
FEATURES
SOURCE
gene
CDS
TITLE
JOURNAL
Gomes, S.L., Gruber, A., Ho, P.L., Hohelsel, J.D., Junqueira, M.L.,
Kemper, E.L., Kitajima, J.P., Krieger, J.E., Kuramae, E.E., Laigret, F.,
Lambert, M.R., Leite, L.C.C., Lemos, E.G.M., Lemos, M.V.F., Lopes, S.A.,
Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M.B.N.,
Madeira, H.M.F., Martins, C.L., Marques, M.V., Martins, E.A.L.,
Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C.,
Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A.,
Nascimento, A.L.T.O., Netto, L.E.S., Nhani, J.A., Nobrega, F.G.,
Nunes, L.R., Oliveira, M.A., de Oliveira, R.C., de Oliveira, R.C.,
Palmeri, D.A., Paris, A., Peixoto, B.R., Pereira, G.A.G., Pereira
Jr., H.A., Pesquero, J.B., Queglio, R.B., Roberto, P.G., Rodrigues, V.,
de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V.,
Sawasaki, H.E., da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva
Jr., W.A., da Silveira, J.F., Silvestri, M.L.Z., Siqueira, M.J., de
Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M.,
Tshako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S.,
Vettore, A.L., Zago, M.A., Zatz, M., Weidman, J., and Setubal, J.C.
Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil

Location/Qualifiers
1..10593
/organism="Xylella fastidiosa"
/db_xref="taxon:2371"
complement(227..1356)
/gene="XF0061"
complement(277..1356)
/gene="XF0061"
/note="similar to SPIP07674 (percent identity: 28 %/query
alignment coverage: 64.9 %/subject alignment coverage:
65.1 %): identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="transcriptional repressor"
/protein_id="AAF82874.1"
/db_xref="GI:9104831"
/translation="MLLIQTLFEIDRATAMSLFGRIRKIPQEOQLTVAFSGEHIE
IDPELIDLEKQMRKDNFGFTTESLMELGMDKQGHPEVILFKNKRGRYLMVG
ERWMLACKAVDRIKAVIRELTDQVRVQRAENIQRENIQLEIATVALDKDRRLCT
LEKVAENKNGIDNINVERINFLVEVETNSLSQAVSGITADITANDLVRLKVNQ
AAKTIIDQAKDPGIMKRIYREKILQNAKPTTNKNAKTTDSRMQESTDQEGETHIQ
NISNOKEPATVASAPLPDPAQALVALAKAQETICTEVSALSKILHQRQVYSALGDV
LQNYQAGALAPRTRHYASTANSK"
complement(1331..1657)
/gene="XF0062"
complement(1331..1657)
/note="hypothetical protein: identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF82875.1"
/db_xref="GI:9104832"
/translation="MYHRSIRORSINHTVKKHCKMSDLQNHIDITPYLPNNTNMI
LEMDNFKNTNEICHTSPISQIYRPPRPHQHNIPFIIPLHPLMPQHLKMYARRA
VVTIDV"
complement(2427..3152)
/gene="XF0063"
complement(2427..3152)
/note="similar to GI15921548 (percent identity: 37 %/query
alignment coverage: 93.4 %/subject alignment coverage:
95.7 %): identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="competence protein F"
/protein_id="AAF82876.1"
/db_xref="GI:9104833"

TITLE Direct Submission
JOURNAL Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Dec 7, 2000 this sequence version replaced 91:8779952.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L10045
Center clone name: 749_1-16

Summary Statistics

Sequencing vector: M13; M77815; 36% of reads
Sequencing vector: Plasmid; n/a; 64% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 185301 bases at least Q40
Consensus quality: 188331 bases at least Q30
Consensus quality: 189713 bases at least Q20

Insert size: 17000; agarose-fp

Quality coverage: 8.3 in Q20 bases; agarose-fp

Quality coverage: 7.7 in Q20.

NOTE: This is a 'working draft' sequence. It currently consists of 25 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 15647: contig of 15647 bp in length
* 15648 15747: gap of 100 bp
* 15748 16321: contig of 574 bp in length
* 16322 16421: gap of 100 bp
* 16422 17197: contig of 776 bp in length
* 17198 17297: gap of 100 bp
* 17298 17904: contig of 607 bp in length
* 17905 18004: gap of 100 bp
* 18005 19904: contig of 1900 bp in length
* 19905 20004: gap of 100 bp
* 20005 21868: contig of 1864 bp in length
* 21869 21968: gap of 100 bp
* 21969 24760: contig of 2792 bp in length
* 24761 24860: gap of 100 bp
* 24861 27006: contig of 2146 bp in length
* 27007 27106: gap of 100 bp
* 27107 29532: contig of 2426 bp in length
* 29533 29632: gap of 100 bp
* 29633 56088: contig of 26456 bp in length
* 56089 56188: gap of 100 bp
* 56189 59962: contig of 3774 bp in length
* 59963 60062: gap of 100 bp
* 60063 65173: contig of 5111 bp in length
* 65174 65273: gap of 100 bp
* 65274 68681: contig of 3408 bp in length
* 68682 68781: gap of 100 bp
* 68782 73500: contig of 4719 bp in length
* 73501 73600: gap of 100 bp
* 73601 78056: contig of 4456 bp in length
* 78057 78156: gap of 100 bp
* 78157 84041: contig of 5885 bp in length
* 84042 84141: gap of 100 bp
* 84142 93337: contig of 9196 bp in length
* 93338 93437: gap of 100 bp
* 93438 102240: contig of 8803 bp in length
* 102241 102340: gap of 100 bp
* 102341 113166: contig of 10826 bp in length
* 113167 113266: gap of 100 bp

```

FEATURES

SOURCE

```

* 113267 123311: contig of 10045 bp in length
* 123312 123411: gap of 100 bp
* 123412 133265: contig of 9854 bp in length
* 133266 133365: gap of 100 bp
* 133366 148098: contig of 14733 bp in length
* 148099 148198: gap of 100 bp
* 148199 165273: contig of 17075 bp in length
* 165274 165373: gap of 100 bp
* 165374 190517: contig of 25144 bp in length
* 190518 190617: gap of 100 bp
* 190618 193351: contig of 2734 bp in length.
* 193351
1. 193351
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-749116"
/clone_lib="RPC1-11 Human Male BAC"
1. 15647
/note="assembly-fragment"
/clone_end:sp6
/vector_side:left"
15748..16321
/note="assembly-fragment"
16422..17197
/note="assembly-fragment"
17298..17904
/note="assembly-fragment"
18005..19904
/note="assembly-fragment"
20005..21868
/note="assembly-fragment"
21969..24760
/note="assembly-fragment"
24861..27006
/note="assembly-fragment"
27107..29532
/note="assembly-fragment"
29633..56088
/note="assembly-fragment"
56189..59962
/note="assembly-fragment"
60063..65173
/note="assembly-fragment"
65274..68681
/note="assembly-fragment"
68782..73500
/note="assembly-fragment"
73601..78056
/note="assembly-fragment"
78157..84041
/note="assembly-fragment"
84142..93337
/note="assembly-fragment"
93438..102240
/note="assembly-fragment"
102341..113166
/note="assembly-fragment"
113267..123311
/note="assembly-fragment"
123412..133265
/note="assembly-fragment"
133366..148098
/note="assembly-fragment"
148199..165273
/note="assembly-fragment"
165374..190517
/note="assembly-fragment"
190618..193351
/note="assembly-fragment"
/clone_end:77
/vector_side:right"

```



```

repeat_region /rpt_family="putative Tourist type MITE element"
complement(13218..13316)
repeat_region /rpt_family="AT-rich"
complement(15045..15100)
repeat_region /rpt_family="AT-rich"
complement(15858..16053)
repeat_region /rpt_family="Wanderer-O11 MITE element from gb:U34601
Oryza sativa wanderer mobile element linked to Xaz1 (411
to 637) 227 nt"
complement(15858..16053)
repeat_region /rpt_family="WANDERER|WANDERER Wanderer mobile element
linked to Xaz1."
complement(15861..16039)
repeat_region /rpt_family="Wanderer-Os2 MITE element from gb:L28995
Oryza sativa 3-hydroxy-3-methylglutaryl coenzyme A
reductase gene (696 to 925) 230 nt"
complement(16306..16507)
repeat_region /rpt_family="WANDERER|WANDERER Wanderer mobile element
linked to Xaz1."
complement(16310..16507)
repeat_region /rpt_family="Wanderer-O11 MITE element from gb:U34601
Oryza sativa wanderer mobile element linked to Xaz1 (411
to 637) 227 nt"
complement(16337..16481)
repeat_region /rpt_family="Wanderer-Os1 MITE element from gb:L28995
Oryza sativa 3-hydroxy-3-methylglutaryl coenzyme A
reductase gene (4476 to 4704) 229 nt"
complement(16642..16708)
repeat_region /rpt_family="AT-rich"
complement(17251..17298)
repeat_region /rpt_family="AT-rich"
complement(20332..20394,21021..21092,21210..21305,
22519..22567,23229..23284,23684..23961,24048..24152,
24280..24388,24567..24700)
gene /gene="OSUNBA0055P24.2"
complement(20332..24700)
/gene="OSUNBA0055P24.2"
note="similar to maize EST A1374558, A1795544"
complement(join(20332..20394,21021..21092,21210..21305,
22519..22567,23229..23284,23684..23961,24048..24152,
24280..24388,24567..24700))
/gene="OSUNBA0055P24.2"
/codon_start=1
/product="unknown protein"
/protein_id="AAG13571.1"
/db_xref="GI:10140739"
/translation="MSTAAPASATTAASLPLRSLSSEPRHVFLEPMVAPVYVARA
RLSAVPAKGGGVAANAASAGSAPGAGDSENPYELISPLDGDQKMAVKRRR
KDAENKDAEHLFLERAYADVMMEQIÖNKRNGVAGSIOVSKDKYADQVPMGP
SLPFLSMASHFYVGEKEGVEGRGIRHMAKRVLRSLGLVLSITLAVSLGTGLANFS

```

```

Query Match 75.2% Score 18.8; DB 12; Length 134058;
Best Local Similarity 90.9%; Pred. No. 2.4e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

CY 4 ggtgcatgcacgacattcgat 25
||||| ||||| ||||| |||||
DB 101294 GGTGCATGCGCAGCATTCGAT 101315

RESULT 22
AE000198 11639 bp DNA BCF 01-DEC-2000
LOCUS DEFINITION Escherichia coli K12 MG1655 section 88 of 400 of the complete
genome.
ACCESSION AE000198 U00096
VERSION AE000198.1 GI:1787189
KEYWORDS
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 11639)

```

```

AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, T., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B., and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
JOURNAL MEDLINE
PUBMED 97426617
9278503
REFERENCE 2 (bases 1 to 11639)
AUTHORS Blattner, F.R.
DIRECT SUBMISSION
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 11639)
REFERENCE 3 (bases 1 to 11639)
AUTHORS Blattner, F.R.
DIRECT SUBMISSION
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 11639)
REFERENCE 4 (bases 1 to 11639)
AUTHORS Plunkett, G. III.
DIRECT SUBMISSION
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 [e-mail: mark@ember.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlin. A public version of the database is accessible at
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). ***The E. coli K12 sequence and
its annotations are periodically updated; this is version MS4. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.
FEATURES
source Location/Qualifiers
1..11639
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
63..89
/note="factor Sigma70; predicted +1 start at 1017661"
86..105
/note="central position to predicted promoter:1"
/bound_moiety="CRP predicted site"
complement(102..133)
/note="factor Sigma70; predicted +1 start at 1017680"
123..575
/gene="ycbG"
/note="ycbG"
123..575
/gene="ycbG"
/function="putative enzyme; Not classified"

```


/note="0150: 100 pct identical to fragment YCBG_ECOLI
 SM:P45569 but has 111 additional N-terminal residues"
 /codon_start=1
 /transl_table=11
 /product="putative dehydrogenase"
 /protein_id="AAC74042.1"
 /db_xref="GI:1787190"
 /translation="MKYQOLENLESGMKMYLVKHKREGELITRYIEASAAQAVDVL
 LSENEPVLVNGWIDIKHNPVLNRMKOTIARRKRHNAEHOTRRKSIDLEFIWQ
 RLAGEAORGRKLTSEITVOLLIEDAENKRYANKKSSLKQDQALLGKE"
 complement(651..1691)
 /gene="ompA"
 /note="b0957"
 /complement(651..1691)
 /gene="ompA"
 /function="membrane; Outer membrane constituents"
 /note="f346: 100 pct identical to OMPA_ECOLI SM: P02934"
 /codon_start=1
 /transl_table=11
 /product="outer membrane protein 3a (11*;G;D)"
 /protein_id="AAC74043.1"
 /db_xref="GI:1787191"
 /translation="MKTAIAIAVALAGFAVAQAARKDNTWTGAKLGMSQYHDTGF
 INNGPIHENOLGAGAFGGYQVNPVVGEMGYDMLGRMPYKGSVENGAAYKAQGVOLTA
 KLGPITDDLDIYRLGGMVRADTKSNVYKGNHDTGSPVPAQGEVETAITPEIATRL
 EYQWNNIGDAHTTGTTPDNGMLSLGYSYRFGCEAAVAPAPAPAEVOTKHFTLK
 SDVLENNKATLKPEGQALDOLYSNLDPKDSGVAVLGCTDRIGSDANQOLISER
 RAOSVVDYLSKGIADKISAKGMESNPVTGNTCDNVKQNALIDCLAPRREIETV
 KGIDVYVQPOA"
 complement(1832..1859)
 /note="factor Sigma70; promoter ompA2; documented +1 at
 1019410"
 /complement(1839..1866)
 /note="factor Sigma70; promoter ompA; documented +1 at
 1019414"
 /complement(2048..2557)
 /gene="sua"
 /note="b0958"
 /complement(2048..2557)
 /gene="sua"
 /function="phenotype: Cell division"
 /note="f169: 100 pct identical to SULA_ECOLI SM: P08846"
 /codon_start=1
 /transl_table=11
 /product="suppressor of lon; inhibits cell division and
 ftsZ ring formation"
 /protein_id="AAC74044.1"
 /db_xref="GI:1787192"
 /translation="MTSGYAHRSSSFSSASKIARVSTENTAGLISEVYREDOPM
 MTOLLPLLOLGQSRMOLWLPKOKLSREWVOASGLPIKMTQISOLSCPHTVES
 MYRALRTGNSVSVIGWLADDLTEEHAEIYDAANGNMGIMRPVSSASHATVQLSG
 LKIHSLNLYH"
 complement(2579..2599)
 /note="central position to sua promoter:14.5"
 /bound_moiety="lexa documented site"
 /complement(2594..2622)
 /note="factor Sigma70; promoter sua; documented +1 at
 1020168"
 /complement(2685..2713)
 /note="factor Sigma70; predicted +1 start at 1020305"
 2776..3405
 /gene="b0959"
 2776..3405
 /gene="b0959"
 /function="orf; Unknown"
 /note="0209: This 209 aa ORF is 31 pct identical (8 gaps)
 to 173 residues of an approx. 224 aa protein TPOX_HAEIN
 SM:P43779"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC74045.1"
 /db_xref="GI:1787193"

/translation="MKSLSYKRIYKSOEVLATIGTIEYKSLFGSYSLIVDDVFAVMS
 DGELYLRACRQSOYCYKHPVWLTYKKGRSVTLTKYVSDSLMNOCKILRLSKYS
 LDAIAEKSRSTRNREIRKDLPMNSFPLEALIGGVGKIDVRAIRIIGAKCMWLRLQON
 SLVTEKILFPLEGACIIIGIHEAALPVARQELAMADSLTRKQEPFAELE"
 complement(3368..5530)
 /gene="b0960"
 /complement(3368..5530)
 /gene="b0960"
 /function="orf; Unknown"
 /note="f720: This 720 aa ORF is 38 pct identical (20 gaps)
 to 692 residues of an approx. 720 aa protein YHF_HAEIN
 SM:P44289"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC74046.1"
 /db_xref="GI:1787194"
 /translation="MAFMLSPILKRYTWSAWLYVARIFALGCTTAPFPMWLDVRLT
 IPLTGVAAALTDLDRLAGRLRLNLTILFCFFIASVLELFPMPWLAIGLTST
 SGFLIGLGIGORYATIAFGALLIATYMLGTSLEYHMOQPMYLAGAVWVNLIG
 HLEPVPRIODNARCVEOLARYLELSKRPDPDIDSOAPLYDALANGILMTLN
 QTKLSLTRRCGRGCRGRTRTLYHYFVADHIERSSSHIQOTIREFRSDVLR
 FORLMSRQGACQQLSRCTILRQYOHDPHERAFTHIDALEMRDNCAPADLTKL
 GFLNLRNLAIDAQLATIESEQALPHNNDENLADSPHGLSDWLRSLRHTPESA
 LFRHVAVMSVLFCFYALIQITGMHNGWILSLSEVCOBNVATRHRLKRIIGTVL
 GIATIGPVLVFEVSLRGOALVLYTVGVLPFAFVNOVYAHATMFTLLVLCFNLDGEG
 FEVALPVIDITLGCALIMAAVASYTPMDQFRLLPMLRATPANCRTYDALIEOHO
 GRNRLATYRIARRDANRRDELA SVSNSSERNVTPQIREAFRLCINHTFTSTIS
 ALGAREQVQTLPELTLAFVLDPAVCYVDALHHPADADERNEALASLKRQMOLEPAD
 SKELPVQVQGLLIALPEIRGLRQRIYGVPOPTPVA"
 complement(5540..5986)
 /gene="ycgF"
 /note="b0961"
 /complement(5540..5986)
 /gene="ycgF"
 /function="orf; Unknown"
 /note="f148: 100 pct to fragment YCGF_ECOLI SM: P37065
 (35aa) but has 113 additional C-terminal residues"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC74047.1"
 /db_xref="GI:1787195"

Query Match 74.48; Score 18.6; DB 1; Length 11639;
 Best Local Similarity 84.08; Pred. No. 3.4e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 gcagtgatcgccagcatctgcat 25
 Db 10623 GCAGCTGATCGCCAGCATGTTGAT 10647

RESULT 23
 AE005286
 LOCUS
 DEFINITION
 Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 110
 of 155.
 ACCESSION
 AE005286 AE005174
 VERSION
 AE005286.1 GI:12514140
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli O157:H7 EDL933.
 Escherichia coli O157:H7 EDL933.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 11641)
 Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
 Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
 Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
 Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoustis,K.,
 Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
 Welch,R.A. and Blattner,F.R.
 Genome sequence of enterohaemorrhagic Escherichia coli O157:H7

JOURNAL Nature 409 (6819), 529-533 (2001)
 MEDLINE 21074935
 PUBMED 11206551
 REFERENCE 2 (bases 1 to 11641)
 AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,

Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
 Postal, G., Hackett, O., Link, S., Boutin, A., Shao, Y., Miller, L.,
 Grodzick, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamoudis, K.,
 Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
 Welch, R.A. and Blattner, F.R.

TITLE Direct Submission
 JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES Location/Qualifiers
 1..11641

source
 /organism="Escherichia coli O157:H7 EDL933"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="Enterohemorrhagic"
 124..576
 /gene="ycbG"
 /note="21306"
 124..576
 /gene="ycbG"
 /function="putative enzyme; Not classified"
 /note="Residues 1 to 150 of 150 are 100.00 pct identical
 to residues 1 to 150 of 150 from Escherichia coli K-12
 strain MG1655: B0956"
 /codon_start=1
 /transl_table=11
 /product="putative dehydrogenase"
 /protein_id="AAG5444.1"
 /db_xref="GI:1251414"

gene
 /translation="MKYQOLENLESGMKWLYKKHREGELITRYEASAAQAVDL
 LSENEPVLVGMIDKHNPELVNRMKQITARRKRHNENHQTTRKSIDLEFIVQ
 RLAGLAQRGRKLTSETIVOLIEDENKEKYANKKSSIKQDQALLGKE"
 complement(652..1692)
 /gene="ompA"
 /note="21307"
 complement(652..1692)
 /gene="ompA"
 complement(652..1692)
 /gene="ompA"
 /function="membrane; Cell envelop: Outer membrane
 constituents"
 /note="Residues 1 to 346 of 346 are 100.00 pct identical
 to residues 1 to 346 of 346 from Escherichia coli K-12
 strain MG1655: B0957"
 /codon_start=1
 /transl_table=11
 /product="outer membrane protein 3a (II*G:d)"
 /protein_id="AAG5444.1"
 /db_xref="GI:12514142"

CDS
 /translation="MKKTATIAVALAGFATVQAQAKNDNTWTGAKLGSQYIDTF
 INNPFHENOAGAGGQVQVYVGFEMGTDLGMRPKGSVENGAAYAAQGVQLA
 KLGPYTDIDDIYTRLGGMWRADTKSNVGNKNDHTSVSPFAGVEYATPELATL
 EYQWNNIGDAHTIGTRPDNGMLSLGVSRGQCEAPVAPAPAPAEQTKFTFLK
 SDVLEFNKATLKEGQALDQLSNLDPKDGSVYVLTDRIGSDAVNOCLEISR
 RASQVVDYLISKIPADKISARGMGESNPVTGCDNVKQRAALDCLADPRVEIV
 KGIDVYTQPOA"
 complement(2049..2558)
 /gene="sua"
 /note="21308"
 complement(2049..2558)
 /gene="sua"
 /function="phenotype: Cell division"
 /note="Residues 1 to 169 of 169 are 100.00 pct identical
 to residues 1 to 169 of 169 from Escherichia coli K-12
 strain MG1655: B0958"
 /codon_start=1
 /transl_table=11
 /product="suppressor of lon; Inhibits cell division and
 ftsZ ring formation"
 /protein_id="AAG5444.1"

gene
 complement(2049..2558)
 /gene="sua"
 /note="21308"
 complement(2049..2558)
 /gene="sua"
 /function="phenotype: Cell division"
 /note="Residues 1 to 169 of 169 are 100.00 pct identical
 to residues 1 to 169 of 169 from Escherichia coli K-12
 strain MG1655: B0958"
 /codon_start=1
 /transl_table=11
 /product="suppressor of lon; Inhibits cell division and
 ftsZ ring formation"
 /protein_id="AAG5444.1"

gene
 /db_xref="GI:12514143"
 /translation="MYTSGYAHRSSTSSAASKIARVSTENTAGLISEVYREDQPM
 MROLPLPLLOCGQOSRMQWMLTPQOKLSREMOVQSGPLTKMQISOLSCHEVS
 MRALRTGNVSIVIGWLADLIEEHAELVDANGNMGMFIMRPVSSASHXRLQSLG
 IKIHSLNLYH"
 2777..3406
 /gene="21309"
 2777..3406
 /gene="21309"
 /function="orf; Unknown function"
 /note="Residues 1 to 209 of 209 are 99.52 pct identical to
 residues 1 to 209 of 209 from Escherichia coli K-12 strain
 MG1655: B0959"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAG5445.1"
 /db_xref="GI:12514144"

CDS
 /translation="MKSLSYKRIYKSQEVYATLGTIEYRSLSFGSYLVDDTVFAMVS
 DEGLYRACEQSAQCVKYPVWLYTKKGRSVTLNRYKVDLSLRNOLKVLRSKYS
 LBAALKEKSTRNTREKLDPMNSFLHLEIIEVGIKQVRLALIGAFKMCWRLRQON
 SLVTEKILFMLEGAILIGHEALPVARROELAEWADSLTPKQEPFAELE"
 complement(3369..5531)
 /gene="21311"
 complement(3369..5531)
 /gene="21311"
 complement(3369..5531)
 /function="orf; Unknown function"
 /note="Residues 1 to 720 of 720 are 99.86 pct identical to
 residues 1 to 720 of 720 from Escherichia coli K-12 strain
 MG1655: B0960"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAG5446.1"
 /db_xref="GI:12514145"

gene
 /translation="MAFMLSPLIKRYTNSAWLYARIFALGCTTAPPMWLGDKLT
 IPTLELVAAALDLDLDRLAGRLNLTILFCFYLASVVELFPMPMFLFAGLST
 SGFTLIGGQRTATFAGALLIATYMTSGSLYEWGQOAPNATLGAQVNVVLLIG
 HLLFVPRLODMNARCVRQOLARVLELTKSMRPPDIEDSOAPLYALANGLMATLN
 QTKSLTLRLRDRQGRTRRLHYFVADIDHERASSSHIYOTLREHFRISDVLFR
 FORLMSQOACQJUSRCILNKPOTODHFERATTHIDALEMRNDGAPADLLKTL
 GFLNLNRAIDQOLATIESEQAOLPHNNXNDELADSDPHGLSDIWLRSRFPESA
 LEFHAARMSLVLCFYGAILIOTIGMHGWIILTSFVQCPNATNRHLKRLITGLV
 GIAIGIPVLWXPSEGLVLLVITGVLEFAPRANVYAHATMFLITLVLCFNLLEG
 FEVALPRVLDITLIGCAIAMAAYSIYMPDQFNPLRMLREATEACRYLDATDEQYH
 GPNRLAYRIARPDANRDARLAVSAYSNMSSPNPTPREFAREAPLILNTPSYIS
 ALCAHREQUTNPETLAFEDDAYCYDDALHHPADBERVNEALASKORMOOLEPRAD
 SKRPLVQGVGLIALLPEIIGRLQRIOTVOPETVSA"
 complement(5541..5987)
 /gene="ycbF"
 /note="21312"
 complement(5541..5987)
 /gene="ycbF"
 complement(5541..5987)
 /gene="ycbF"
 /function="orf; Unknown function"
 /note="Residues 1 to 148 of 148 are 99.32 pct identical to
 residues 1 to 148 of 148 from Escherichia coli K-12 strain
 MG1655: B0961"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAG5447.1"
 /db_xref="GI:12514146"

CDS
 /translation="MRTVNLNINFLVYGAFATLGLWLATVSIATLFTPLTRSCWEI
 TKLSVYGENEAIHVELNIPACKRVLLNGLGVNLINFWLIFPGMMLCIAMHATGIAOC
 ISIGIPVCIANFKTIAIALIMPVGRVSVYEAQAARENARRE"
 6110..8164
 /gene="hcd"
 /note="21313"
 6110..8164
 /gene="hcd"
 /function="enzyme; Macromolecule synthesis, modification:
 DNA - replication, repair, restr./modifc'n"
 /note="Residues 1 to 684 of 684 are 99.85 pct identical to

AUTHORS
TITLE
JOURNAL

Mori, H.
Direct Submission
Submitted (29-JUL-1996) to the DDBJ/EMBL/GenBank databases.
Hirotada Mori, NARA Institute of Science and Technology, Res. &
Edu. Center for Genetic Info.: 8916-5 Takayama, Ikoma, Nara 630-01
Japan (E-mail:hmori@etc.aist-nara.ac.jp, Tel:81-7437-2-5660,
Fax:81-7437-2-5669)

COMMENT

gene

CDS

EYOWMNNIGAAHTTKEGPAADMLISYVRFPDGGEEAIVVAAPAPAEVOTKHFHTLK
 SDVAFNRKAKITKEGGAADMLDOLISLNSLMDPDGSSVYLGTDRISGDADNOCLESR
 RAQSVVDVILISKIGIPDAKISARGMESNPVTGNTCDNKKORALIDCLAPRRRIEIVR
 KGIDKVVYVQPOGA"
 complement(1933, .2442)
 /gene="gfiA"
 /note="gula"
 complement(1933, .2442)

CDS

CDS

FEATURES

SOURCE

CDS

CDS

CDS

CDS

/product="Helicase (EC 3.6.1.1) IV."
 /protein_id="BA35727.1"
 /db_xref="GI:1651471"
 /translation="MELKATTLGKRLAHPYDRAVILNAGIKVSGDRHEYLIPNOL
 LKHCRLGLWELEFLVLPDEKVLRLHTEGEGTQFHHLDHMRMGSELSASGV
 LKQDLRLATRTGKMLTREGTSGVOQIROLALPLPVRNLEEFNCEARCKO
 AMLKIESARLONQAYTEAMLETEDFROVESPLPAPARRVNEHSLYLACG
 GSKTSLVARAGMLARGFASPEQIILLAGRKAREPDERIRERLRETTARTH
 ALAHIIIOGSKPIYKILINDIARHEIFTAERKOCSEKKAQAKOMRWLTEMQ
 MSVPEGNWDEKQRLASRLUDRWSLMRHGGQAEMIASAPELRDLFSKRIKL
 APLKAMGALKENAVDSGLIHOAVILEKGRISPMKHILVDEQDITSPORAIL
 AALRKONSQTLTFAVDQWQAIYRPSGAKSLTTFAPENFEGECRDLITRNSRI
 GEVANRFIOQNGOLKPLNSLTNGDKKAVLLDESDOLDLTLGAKPEEILILI
 ARYHMRPASLEKATRPWKLOIDMTTHASKGOADYVILVIGIOESDGPAAARS
 IMEALLPPEVDEPDABERRRLMYALTARIRVALLPKENPFPVELKMLDVPAR
 KP"
 complement(8080,.8547)
 /gene="yccg"
 complement(8080,.8547)
 /gene="yccg"
 /note="ORF ID:0223#8
 similar to SwissProt Accession Number P37066"

Query Match 74.4%; Score 18.6; DB 2; Length 13832;
 Best Local Similarity 84.0%; Pred. No. 3.4e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcaggtgcatgcgcagcatlccgat 25
 ||||| || ||||| ||||| |||||
 Db 10508 GCAGGTGATGCCAGCATTTTGAT 10532

RESULT 26
 AC026813
 LOCUS
 DEFINITION Mus musculus chromosome 5 clone RP23-119M19 strain C57BL6/J,
 WORKING DRAFT SEQUENCE, 15 unordered pieces.
 AC026813
 VERSION AC026813.1 GI:7321458
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 200631)
 Beckstrom-Sternberg,S.M., Benjamin,B., Bouffard,G.G.,
 Dietrich,N.L., Eagle,W.O., Gan,W., Gupta,J., Ho,S.-T., Huang,M.C.,
 McCloskey,J.C., Morse,E., Ojodu,M.A., Pearson,R., Stantripop,S.,
 Summers,T.J., Thomas,J.W., Thomas,P.J., Tlonsong,E.E.,
 Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A., Wetherby,K.D. and
 Green,E.D.
 NISC Mouse Sequencing Initiative
 Unpublished
 2 (bases 1 to 200631)
 Green,E.D.
 Direct Submission
 Submitted (24-MAR-2000) NIH Intramural Sequencing Center, 8717
 Grovemont Circle, Gaithersburg, MD 20877, USA
 ----- Genome Center
 Center: NIH Intramural Sequencing Center
 Center code: NISC
 Web site: http://www.nisc.nih.gov
 Contact: nisc-mouse@hgrl.nih.gov
 ----- Project Information
 Center project name: xh
 Center clone name: 119M19
 ----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 193127 bases at least Q40
 Consensus quality: 195226 bases at least Q20

Consensus quality: 196033 bases at least Q20
 Insert size: 197000; agarose-fp
 Insert size: 192000; pulse-field-gel
 Insert size: 200631; sum-of-contigs
 Quality coverage: 7.35x in Q20 bases; agarose-fp
 Quality coverage: 7.55x in Q20 bases; pulse-field-gel
 Quality coverage: 7.23x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 15 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2735: contig of 2735 bp in length
 * 2736 5031: gap of unknown length
 * 5032 7390: gap of unknown length
 * 7391 10254: gap of unknown length
 * 10255 14682: gap of unknown length
 * 14683 18489: gap of unknown length
 * 18490 25669: gap of unknown length
 * 25670 38267: gap of unknown length
 * 38268 47011: gap of unknown length
 * 47012 60788: gap of unknown length
 * 60789 75761: gap of unknown length
 * 75762 89840: gap of unknown length
 * 89841 117870: gap of unknown length
 * 117871 160484: gap of unknown length
 * 160485 200631: contig of 40147 bp in length.
 FEATURES
 source
 1..200631
 /organism="Mus musculus"
 /strain="C57BL6/J"
 /db_xref="taxon:10090"
 /chromosome="5"
 /clone="RP23-119M19"
 /clone_lib="RECT mouse BAC library 23"
 BASE COUNT 59764 a 40683 c 40100 g 60060 t 24 others
 ORIGIN

Query Match 74.4%; Score 18.6; DB 70; Length 200631;
 Best Local Similarity 84.0%; Pred. No. 2.9e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcaggtgcatgcgcagcatlccgat 25
 ||||| ||||| ||||| |||||
 Db 104431 GCAGGTGATGCCAGCATTTTCAT 104455

RESULT 27
 AC068663
 LOCUS
 DEFINITION Mus musculus chromosome 5 clone RP23-280D18 strain C57BL6/J,
 WORKING DRAFT SEQUENCE, 25 unordered pieces.
 AC068663
 VERSION AC068663.2 GI:8134859

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 212908)
AUTHORS Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W.,
Bonifard, G.G., Dietrich, N.L., Eagle, W.O., Gupta, O., Ho, S.-L.,
Huang, M.C., Idol, J., Lee-Edin, S.-Q., Maduro, O.L., Maduro, V.B.,
Mastrian, S.D., McCloskey, J.C., Morse, E., Ojodu, M.A., Pearson, R.,
Stantiprop, S., Summers, T.J., Thomas, J.W., Thomas, P.J.,
Tingson, E.E., Touchman, J.W., Tran, J.T., Vogt, J.L., Walker, M.A.,
Wetherby, K.D., and Green, E.D.
NISC Mouse Sequencing Initiative
Unpublished
2 (bases 1 to 212908)
Green, E.D.
Direct Submission
Submitted (06-MAY-2000) NIH Intramural Sequencing Center, 8717
Groomet Circle, Gaithersburg, MD 20877, USA
On Jun 1, 2000 this sequence version replaced gi:7712112.

Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nigr.nih.gov

Project Information
Center project name: xl
Center clone name: 280D18

Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 199975 bases at least Q40
Consensus quality: 202777 bases at least Q30
Consensus quality: 204027 bases at least Q20
Insert size: 222000; agarose-gel
Insert size: 216000; pulse-field-gel
Insert size: 210508; sum-of-ctrls
Quality coverage: 4.37x in Q20 bases; agarose-gel
Quality coverage: 4.49x in Q20 bases; pulse-field-gel
Quality coverage: 4.61x in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2371: contig of 2371 bp in length
*
2372 2471: gap of unknown length
*
2472 4622: contig of 2151 bp in length
*
4623 4722: gap of unknown length
*
4723 7916: contig of 3194 bp in length
*
7917 8016: gap of unknown length
*
8017 10417: contig of 2401 bp in length
*
10418 10517: gap of unknown length
*
10518 14260: contig of 3743 bp in length
*
14261 14360: gap of unknown length
*
14361 18020: contig of 3660 bp in length
*
18021 18120: gap of unknown length
*
18121 22445: contig of 4325 bp in length
*
22446 22545: gap of unknown length
*
22546 26537: contig of 3992 bp in length
*
26538 30194: gap of unknown length
*
30195 30284: gap of unknown length
*
30295 34805: contig of 4511 bp in length
*
34806 39674: gap of unknown length
*
39675 39774: gap of unknown length
*

FEATURES
source
* 39775 44570: contig of 4796 bp in length
* 44571 44670: gap of unknown length
* 44671 49669: contig of 5299 bp in length
* 49670 50069: gap of unknown length
* 50070 55625: contig of 5556 bp in length
* 55626 55725: gap of unknown length
* 55726 61502: contig of 5777 bp in length
* 61503 61602: gap of unknown length
* 61603 70932: contig of 9329 bp in length
* 70932 79217: contig of 8186 bp in length
* 79218 79317: gap of unknown length
* 79318 88380: contig of 9063 bp in length
* 88381 88480: gap of unknown length
* 88481 102624: contig of 1414 bp in length
* 102625 102724: gap of unknown length
* 102725 112658: contig of 9934 bp in length
* 112659 112758: gap of unknown length
* 112759 130169: contig of 17411 bp in length
* 130170 130269: gap of unknown length
* 130270 149318: contig of 19049 bp in length
* 149319 149418: gap of unknown length
* 149419 170076: contig of 20658 bp in length
* 170077 170176: gap of unknown length
* 170177 189732: contig of 19556 bp in length
* 189733 189832: gap of unknown length
* 189833 212908: contig of 23076 bp in length.

Location/Qualifiers
1. 212908
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone="RP23-280D18"
/clone_1b="RPC1 mouse BAC library 23"
1. 2371
/note="assembly-fragment"
2472. 4622
/note="assembly-fragment"
4723. 7916
/note="assembly-fragment"
8017. 10417
/note="assembly-fragment"
clone_end:SP6
vector_size:right"
10518. 14260
/note="assembly-fragment"
14361. 18020
/note="assembly-fragment"
18121. 22445
/note="assembly-fragment"
22546. 26537
/note="assembly-fragment"
26638. 30194
/note="assembly-fragment"
30295. 34805
/note="assembly-fragment"
34906. 39674
/note="assembly-fragment"
39775. 44570
/note="assembly-fragment"
44671. 49669
/note="assembly-fragment"
50070. 55625
/note="assembly-fragment"
55726. 61502
/note="assembly-fragment"
61603. 70931
/note="assembly-fragment"
71032. 79217
/note="assembly-fragment"
79318. 88380
/note="assembly-fragment"

```

misc_feature      88481..102624
                    /note="assembly_fragment"
misc_feature      102725..112638
                    /note="assembly_fragment"
misc_feature      112759..130169
                    /note="assembly_fragment"
misc_feature      130270..149318
                    /note="assembly_fragment"
misc_feature      149419..170076
                    /note="assembly_fragment"
misc_feature      170177..189732
                    /note="assembly_fragment"
misc_feature      189833..212908
                    /note="assembly_fragment"
BASE COUNT      61004 a 43094 c 43404 g 62838 t 2568 others
ORIGIN
Query Match      74.4%; Score 18.6; DB 73; Length 212908;
Best Local Similarity 84.0%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 gcaggtgcacgcgcacattcgat 25
      ||||| ||||| ||||| ||
Db 145018 GCAGGTGCATGCATGATTTCAT 145042

```

```

RESULT 28
AP002554 327773 bp DNA BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 5/20.
ACCESSION AP002554 BA000007
VERSION AP002554.1 GI:13360491
KEYWORDS
SOURCE
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:R1MD 0509952)
DNA.

```

```

ORGANISM
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (sites)
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C., and Shinagawa, H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
2 (sites)
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H., and
Hayashi, T.
Comparative analysis of the whole set of RNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356

```

```

TITLE
JOURNAL MEDLINE
REFERENCE
AUTHORS
Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C., and
Shinagawa, H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M., and

```

```

TITLE
JOURNAL MEDLINE
REFERENCE
AUTHORS
Shinagawa, H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 327773)
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H., and
Hayashi, T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken-gen-info.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
Fax: 81-6-6879-2047)
genome project.

```

```

COMMENT
FEATURES
SOURCE
location/Qualifiers
1..327773
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="R1MD 0509952"
/db_xref="taxon:83334"
150..2258
/gene="ECS1032"
150..2258
/gene="ECS1032"
/note="similar to YCBY_ECOLI g111787182 percent identity
99 in 702 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="putative oxidoreductase"
/protein_id="BAB34455.1"
/db_xref="GI:13360492"
/translation="MNSLPASTARLEBELKTEENLGAVECQVQVGVHFKGDTRLV
YQSLMWSRLASRLMPLGECXYSDLDLYLGOALNMTENPAGTFAVHSGNDIT
RNSQYGAAMKVDIAIDAFTRKMLPPNPVNDAPDIPRVNVMILKETAISALDSGGLH
LNGYDRAGIAIKETLAAAIYVRGMPGPPPLDPCGSGGLILEAMLMATDRAPGL
HRRMGFGSQAQDEHAIQOEYKAEQOTARGKLAEYSHFPGSDAVIQRATNAR
LAGLELITFEFYKDYAOITNPLKPPYCTYSNPYGRDSEPRLLIALHSLGRIMK
NPGGWNLSFSASPDLSQDLRADKOTKKANGLUDCVQNHVAESTPUSKPMVA
EYANRKRNLKFEKRWARQESIECYRYLDLDPYNAVDVRYADWVVQVQAPKTI
DAHKARQRFEDIIAATISVLGIAPKLVLTREKQKNOYOKLGEKQFLEVYNA
HLMVNLTYLDLTGLDRIARMLGOMSKGKIDLSYGSATVHAGLGGARSTT
HMSRSTYLEMARNLRLNGLTGRAHRLIOACILAMLRANQOPDILFDPTFNSKR
MEDADPVORHIALMKDLKRLIRAGSTIMFSNNRGRFMDLGLAKIGKAKQETTKT
ISODPARROTHCMCLITAA"
2270..4177
/gene="ECS1033"
2270..4177
/gene="ECS1033"
/note="similar to UUP_ECOLI g11787182 percent identity 99
in 635 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="putative ATP-binding component of a transport
system"
/protein_id="BAB34456.1"
/db_xref="GI:13360493"
/translation="MSLSMGAWSFSDAPLIDNNAELHEDNRVCLVGRNGAKST
LMKLTNRQGLDGGRTIYEDLIVARLQDDPRPVEGSDYFVAGIEDEAEYLRKH
DISRLVMDPSEKINLNLAKVQEDLDHINIMQENRINEVLAQLGDPNVALSLSGQ
MLRKALGRLAVSNPRVLLDEPNHLDIEITIMLEGFLKFNCTIIFISIDRFIKR
MATRIYDIDRGLVTVYDGNVYLLKEALRVELEONAEFRDLAQDEWIRGQIKR
BRTNRGVRALKMRRRGRRRVMGTAKQVEASRSQTVPEMDQVYQNGKQ
VKDSQVLRDRTKALICPNCCTTLIKLMDGLOLQDSGRIHGTLELVAYFDQHA
ELDPKTYMDNLASGQEVNNGKPRHYLGLTLOFLHPKRAMTPVALSGGSENRLL
LARFLKPSNLLIDEPNDLIDVLELLELIDSYOGTVLASHDQFVNDVTBECW
IFEQGGKIRGVGYHDARGQEQVYALAKPAAVKNNEPAKAEYIKRSSSKYSTKL
QRELEQILPOLLEDEAKLEALQTOVADAFSSQPHQOTQVILADMAAAEDELQAFER
MEYLEALKNG"
4307..5560
/gene="ECS1034"

```

gene


```

CDS
4307..5560
/gene="ECS1034"
/Note="similar to POIA_ECOLI g111787183 percent identity
99 in 417 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not-experimental
/product="paraquat-inducible protein A"
/protein_id="BAB34457.1"
/db_xref="GI:13360494"
/translation="MCEHHAAKHIICSGCDMLVALPRLHSGOKAACPCGCTTLTVAM
DAAPROPTAVYALALPMLLSLFPVNNVAGVSEITLLEIPGLSEEDVAGTGE
FILFQVYAPFCLITLITLNNRDELPPVKNKOLAVRPLKTMGAELFACVAVFV
KLMATGSGVGSFPLPWCFCVQLRAFCVCVRNMDIAPPELROPKRGVYGR
OGILRSCGCTALLPADEPVCPCGKRGYRRNSLQMTLALLVTSIMLYLPANLPI
VNDLGSCKMPTILLAGVILLMSGSYPVAIVFLASIMPTLKMALIAWLCWDAGHG
KRDERMHLIYEVEFVGRMSMIDVFIIVIALVLMGGLMSITPAMGLMALVAVIM
TMPSAMTFDPRLSMRQPESEHES"
5565..7205
/gene="ECS1035"
5565..7205
/gene="ECS1035"
/Note="similar to POIB_ECOLI g111787184 percent identity
99 in 546 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not-experimental
/product="paraquat-inducible protein B"
/protein_id="BAB34458.1"
/db_xref="GI:13360495"
/translation="MESNNGEAKIOLKWKMSPPVFIPIYALIGAVLPHYHSGQPE
VTLITNABETGEGKTTIKRSYDVGVESATLADLTVEIKARLNSMELIKHDT
VFWVVRPOIGREGISGLGTLISGVYELDPPGASGMDYDLDSPLAPPAKGIKV
VLDSKAGQLSPEDPVLFRGVRGVSSTFDPOKRNISVOLFNPVRLVITSNRF
WKDSGIAVDLSAGMRVEMSGITLTLSCGSEFDPGLDQGVAPKTAFLVYDDKS
IODSLYDHDIDYLMFKDSYRGDPCGAPFERRIGLTVSKVFPAPNRRQFENDYR
IPVLIIEPRILKMOGENDAVDEHIGELIKGSLKNGNLVTGALVVDDEYNT
PATGTIRENGYOIIPVSGLAGLQORRLKALDKINKLPINMIGQASTISEORT
MKNLQTLDSMKNILASOSMOQLPTMOSTLRLNSMGOFGOSAAINMADMRQL
DOVLELPVLTILNKSNAIVEAKDKDPEPRAKQ"
7217..7765
/gene="ECS1036"
7217..7765
/gene="ECS1036"
/Note="similar to YWBA_ECOLI g111787185 percent identity
98 in 182 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not-experimental
/product="hypothetical protein"
/protein_id="BAB34459.1"
/db_xref="GI:13360496"
/translation="MTIALMLAGSCSGEINKNYQLPVVSGTOSTASGRNLLWE
OVANPDYLAGNVYVSDVKYIANNMLASPDQDLRLTLVANNLSTOLPGWVVASQ
PLGSAOPTLVATVTEFNGRYDKGIYSGEMLLNHGOGLIRPRRLSEVYQDDYDEV
KVLGVMSQERASIADEIKRLP"
8021..8188
/gene="ECS1037"
8021..8188
/gene="ECS1037"
/Note="identical to RMF_ECOLI g111787186 (Conserved in
E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not-experimental
/product="ribosome modulation factor"
/protein_id="BAB34460.1"
/db_xref="GI:13360497"
/translation="MKROKRDRLERAHQRGYQAGIACRSKEMCPYOTLNORSQWLGM
REAMADRYVMA"
8258..8776
/gene="ECS1038"
8258..8776
/gene="ECS1038"
/Note="similar to FABA_ECOLI g111787187 percent identity
100 in 172 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not-experimental
/product="beta-hydroxydecanoyl thioester dehydratase"
/protein_id="BAB34461.1"
/db_xref="GI:13360498"
/translation="MVDKRESYTTEEDLLASGRPLFGAKGQPLPAPNMLMDRVYKMT

```

```

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 2; Length 327773;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db 21291 GCAGCTGAATGCCAGCATGTGAT 21315

```

```

RESULT 29
AP002998 347750 bp DNA BCT 03-FEB-2001
LOCUS Mesorhizobium loti DNA, complete genome, section 5/21, complete
DEFINITION sequence.
ACCESSION AP002998 BA000012
VERSION AP002998.1 GI:11994973
KEYWORDS HTG.
SOURCE Mesorhizobium loti (strain:MAFF303099) DNA.
ORGANISM Mesorhizobium loti
Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.
1 (sites)

```

```

REFERENCE
AUTHORS Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,
Watanabe, A., Idegawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,
Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,
Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpou, S., Sugimoto, M.,
Takeuchi, C., Yamada, M. and Tabata, S.
Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
JOURNAL DNA Res. 7 (6), 331-338 (2000)
MEDLINE 21082930
REFERENCE 2 (bases 1 to 347750)
AUTHORS Kaneko, T.
TITLE Direct Submission
Institute (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp,
URL:http://www.kazusa.or.jp/rhizobase/, Tel:81-438-52-3935,
Fax:81-438-52-3934)

```

```

FEATURES
source 1..347750
Location/Qualifiers
/organism="Mesorhizobium loti"
/strain="MAFF303099"
/db_xref="taxon:381"
BASE COUNT 63406 a 113924 c 108801 g 61619 t
ORIGIN

```

```

Query Match
Best Local Similarity 74.4%; Score 18.6; DB 2; Length 347750;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Db 165017 GCAGTGCACACACGCTTTCAT 165041

```

```

RESULT 30
AC013236 119797 bp DNA HTG 03-NOV-1999
LOCUS AC013236
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered

```

```

pieces.
ACCESSION AC013236
VERSION AC013236.1 GI:6223096
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 119797)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (03-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10214372 by the submitter.
For further information on this sequence you may e-mail to
fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Source Location/Qualifiers
1..119797
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 35319 a 25050 c 25201 g 34227 t
ORIGIN
Query Match 73.6%; Score 18.4; DB 63; Length 119797;
Best Local Similarity 95.0%; Pred. No. 3.7e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 6 tgcattgcagcatttcatt 25
||||||| |||||||
DB 17322 TGCATCGCTGCATTTCGAT 17303

RESULT 31
AE003644 264853 bp DNA INV 04-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 142000013386055 section 37
ACCESSION AE003644 AE002690
VERSION AE003644.2 GI:10728786
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 264853)
AUTHORS Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazek,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Fiankuch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brotlier,P., Burlis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dobson,K., Doup,I.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Foster,C., Gabriellian,A.E., Gary,N.S.,
Geisbert,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Mel,M.H., Idegian,C., Jaitani,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Mishina,N.V., Mody,C.,
Morris,J., Moshell,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacle,J.M., Palazolo,M., Platten,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Sinden,Kamos,I.,
Stapleton,M., Strong,R., Sun,E., Svitek,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhang,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
2 (bases 1 to 264853)
JOURNAL Science 287 (5461), 2185-2195 (2000)
MEDLINE 20196006
REFERENCE 20196006
AUTHORS Adams,M.D., Celisner,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT On Oct 9, 2000 this sequence version replaced gl:7298162.
FEATURES
Source Location/Qualifiers
1..264853
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
Joln(20111..20584,20887..23268)
/gene="noc"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT14619"
/db_xref="FLYBASE:FBgn0005771"
/db_xref="FLYBASE:FBgn0005771"
<20111..>23268
/gene="noc"
/note="CG4491"
/map="35A4-35B1"
/db_xref="FLYBASE:FBan0004491"
/db_xref="FLYBASE:FBgn0005771"
Joln(20495..20584,20887..22410)
/gene="noc"
/note="noc gene product; Nucleotide sequence of the Celera
sequence differs from the published sequence for this
transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0004491"
/db_xref="FLYBASE:FBgn0005771"
/protein_id="AAFS339.1"
/db_xref="GI:7298163"
/translacion="MVFLEGGGCVTVIGNNOYLPDYIAPITPTMDAKKSPALLAOT
CSOIGADSSAVPLIAMDKNKTKPCACSSSSSSSSGSAISAKSPGOGAKSPKS
TPISATSASISNTSTGEITKAEKPYPTVNTSHONONSFKSSILDAEPRPSKNS
SSAORRVRSKSNATPPDGGKAEISANDSSSKRTYSPSSSGRGSAPYRSMEVL
NNANTTOHPKREMSMAAANAANAATAAATPAGVIGLNPUSALCCPGEQGANPAFRP
FAGGSHHAAMLAVANGGYPGAPEGGAGOPNPYISYRITTPAGGEAIVPCD
PYGDCPYSAHTLMGACPCACGCTDEHQYGLGAMASAGLPPAPYSGAANAANA
NAAARSAPYCSWVGADYCGKRFOTDELSHLRTHTGNLSPPAANAALASOAK
SLIGLTFPSALRAGYPTPLSPSSAANAARHPYAKPPGALAGSPRPGAGARN
PAAAAAALCPYSPYPMQGRGAHQ"
complement(Joln(76890..77044,77116..77503,77900..>78254))
/gene="Bg:DS04641.8"
/product="CT11693"
/db_xref="FLYBASE:FBan0004218"
/db_xref="FLYBASE:FBgn0028881"
complement(<76890..>78254)
/gene="Bg:DS04641.8"
/note="CG4218"
/map="35B1-35B2"

```

Query Match	Best Local Similarity	95.0%;	Pred. No. 3.5e+02;	Matches	19;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
Oy	6	tgcagccagcattcgat	25										
Db	100785	TGCATCCCTGCATWTCGAT	100804										
RESULT 32													
DROSADH04													
LOCUS													
DEFINITION													
ACCESSION													
DROSADH04	300962 bp	DNA	INV										
Drosophila melanogaster	chromosome 2L, region 34C4-36a7	(Adh	21-MAR-2000										
region, section 4 of 10 of the complete sequence.													
AE003410	AC001647	AC000044	AC2000421	AC000420	AC000423	AC000602							
AC000422	AC000426	AC001663	L77007	L77005	L77008	L77006							
AC000603	AC000604	AC000605	AC000606	AC001665	L81441	L81442	L81443						
AC002440	L79841	L79849	L81431	L79843	L79845	AC000579	L79844	L79846					
AC001794	L79852	AC000580	L79839	L79848	L79851	L79851	AC000581						
AC000349	AC000582	L79842	AC000583	L81429	L79855	L81428	L79853						
AC000584	L81330	L79847	L79850	L81430	AC000585	AC000586	L79840						
AC004118	L42034	L42035	L39710	L39711	L39712	L39713	L39714						
L39716	L39717	AC000576	AC001793	AC000575	L34443	L43443	L43445						
L43446	L43447	L46823	L76993	L76994	L76995	L42036	L42037	L42038					
L39718	L39719	L39720	L39721	AC004359	L49192	L49193	L79860	L79861					
L79862	AC000591	L79863	AC000587	L76996	L76997	L48466	AC000589						
AC000588	AC000590	L43460	L43461	L43462	L43463	L43464	L43465	L43466					
L43467	L43468	L43469	L43470	L43471	L49403	L35753	L35736	L35744					
L35738	L35758	L35747	L35742	L35757	L35741	L35749	L35755	L35748					
L39673	L35761	L35759	L35753	L35740	L35746	L35751	L35757	L35752					
L35745	L35756	L35760	L35762	L35750	L35743	L35739	L35754	L49404					
L39681	L39678	L39677	L39680	L39675	L39674	L39679	L39676	AE003406					
AE003410.1	GI:7287571												
HTG.													
4 of 10													
fruit fly.													
Drosophila melanogaster													
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;													
Phylogoria; Neoptera; Endopterygota; Diptera; Brachycera;													
Muscomorpha; Ephyritidae; Drosophilidae; Drosophila.													
1 (bases 1 to 300962)													
Ashburner, M., Mista, S., Roote, J., Lewis, S.E., Blazer, R., Davis, T.,													
Doyle, C., Calle, R., George, R., Harris, N., Hartzell, G., Harvey, D.,													
Hong, L., Houston, K., Hoskins, R., Johnson, G., Martin, C.,													
Moshrefi, A., Palazolo, M., Reese, M.G., Spradling, A., Tsang, G.,													
Man, K., Whitehead, K., Celisner, S. and Rubin, G.M.													

REFERENCE
AUTHORS

2 (bases 1 to 300962)
Celisner, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
Butenhorff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farlan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Sult, E.,
Svilaras, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rudin, G.M.

TITLE
JOURNAL

Direct Submission
Submitted (08-MAR-2000) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, MS 64-121, Berkeley, CA
94720, USA

COMMENT

On Mar 22, 2000 this sequence version replaced gi:2337893
gi:1945577 gi:3006208 gi:2822114 gi:1945595 gi:2342706.
Submitted by the Berkeley Drosophila Genome Project. For more
information, visit the BDCP Web site: <http://www.fruitfly.org/> This
is the finished sequence of 34C4-36A7.
The orientation of this sequence along the chromosome is left to
right. This sequence was annotated by Sima Misra
(simam@fruitfly.berkeley.edu) on behalf of the Berkeley Drosophila
Genome Center. Coding sequences are predicted based on
computational analysis, using both gene and CDS prediction programs
and matches to other sequences. These predictions and matches have
been evaluated by the annotators and may have been refined by hand.
The annotators have also used their judgement about which matches
to include in this record. The annotations on this sequence can be
examined in more detail from
<http://www.fruitfly.org/publications/Adh.html>
The annotation syntax used in this record is documented at
ftp://ftp.ebi.ac.uk/pub/databases/edg/sequence_annotation_README.v1.2.

FEATURES
SOURCE

1. 300962
Location/Qualifiers
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="34C4-36A7"
/clone="P1 DS01368 (D100), P1 DS08249 (D25), P1 DS08284
(D247), P1 DS00941 (D28), P1 DS08220 (D39), P1 DS00180
(D29), P1 DS01514 (D33), P1 DS00131 (D35), P1 DS05819
(D22), P1 DS01759 (D41), P1 DS01523 (D34), P1 DS01652
(D43), P1 DS03792 (D31), P1 DS01068 (D37), P1 DS06238
(D26), P1 DS08340 (D38), P1 DS04641 (D30), P1 DS01160
(D42), P1 DS01486 (D32), P1 DS09219 (D101), P1 DS07721
(D20), P1 DS00810 (D21), P1 DS06874 (D19), P1 DS03431
(D102), P1 DS03144 (D106), P1 DS03323 (D127), P1 DS01219
(D129), P1 DS00929 (D48), P1 DS04929 (D104), P1 DS03192
(D103), P1 DS09194 (D267), P1 DS07295 (D47), P1 DS05639
(D11), P1 DS07851 (D49), P1 DS03023 (D8), P1 DS01845
(D7), P1 DS04862 (D107), P1 DS07108 (D51), P1 DS09217
(D109), P1 DS02252 (D97), P1 DS00365 (D108), P1 DS07486
(D46), P1 DS00913 (D24), P1 DS04095 (D112), P1 DS02795
(D50), P1 DS07473 (D113), P1 DS02740 (D27), P1 DS09218
(D114), P1 DS02780 (D23), BAC BACR48E02 (D482), BAC
BACR44L22 (D483)"
/clone_lib="P1 library, partial Sau3A in pNS582ctcttAD10
and RRC1-98 (Roswell Park Cancer Institute Drosophila
melanogaster BAC library, partial EcoRI in pBACe3.6)"
complement(11154..14637)
/gene="BG:DS08340.1"
complement(join(11154..11215,11685..11770,12367..12651,
13069..13190,14272..14637))
/gene="BG:DS08340.1"
/note="symbol=BG:DS08340.1; prediction-method='genscan',
version='1.0'; Date run: 25-Jan-99; Time: 18:30:10',
score='41.800'; match-method='BLASTN',
version:'2.0a19mp-Washu [05-Feb-1998] [Build sol2.5-ultra
GH Drosophila melanogaster head port2 Drosophila
melanogaster cDNA clone GH12842 5prime, mRNA

sequence: A1135185"
/codon_start=1
/product="hypochothelial protein"
/protein_id="AAFA4876.1"
/db_xref="GI:7287838"
/db_xref="FLYBASE:FBgn0004857"
/db_xref="FLYBASE:FBgn0989943"
/translation="MIRSCRALISQGLRLRCSLAKAASNLKDDVEEMETPANRQ
VELPNNPEKLSKRYLAFREKLRSEAPLEPCAPHPAHEKEPLKPPNNTPYTG
ICGOAGPEPTRGDMERKGRKSTFOIPIDNLANKOQAREATTSIIHOMPSNDVG
EGRMNAPDHOSCRAYEISRSRKYLELMERKKEAVNDOPFTHLRDFAFASISYWRK
TSTDKLEIAIACGETTRQSSIGETISGKHFTMSKETTECVLSERTITITITLTP
VWRGSGGATFPFPAIVHFRSRPKNG"
complement(29417..32836)
/note="Doc element: DS08340.297"
/function="transposon"
/db_xref="FLYBASE:FBgn0000481"
join(54637..55109,55412..57793)
/gene="noc"
/gene="noc"
54637..57793
join(55020..55109,55412..56935)
/gene="noc"
/note="symbol=noc; synonymy=BG:DS04641.1;
match-method='sim4'; score='1000.0';
desc:'Genbank::L14009:Drosophila melanogaster zinc finger
protein (nocA) mRNA. CDS:385..1998; PID:9457883.'"
species:'Drosophila melanogaster';
match-method:'BLASTX'; version:'2.0a19mp-Washu
[05-Feb-1998] [Build sol2.5-ultra 01:47:30 05-Feb-1998]';
score:'167.0'; desc:'trembl:090237:VITELLOGENIN
(PRAMENT); organism:ANOLIS PULCHELLUS. dbxref:Genbank:
U46857; gi197667; ', species:'ANOLIS PULCHELLUS.'"
/codon_start=1
/evidence=experimental
/protein_id="AAFA4877.1"
/db_xref="GI:7287839"
/db_xref="FLYBASE:FBgn0005771"
/db_xref="FLYBASE:FBgn0989942"
/translation="WVLEGGGVVTTGNNOYLPDIAPLPTTMDAKKSPALLAQT
CSQIGADSSAVKPLIAMDKNTKPKACSSSSSSSSGSAISAKSPGQAKSPKSS
TPISTATASISNTSTGEIKLAFKPYTNVONONSPKSSSLDAEPTRPSSKKS
SSAQRVRSKSNATPTDGGKAEISADSSSSSKRYSPSSGSRGASPIYRSMEVL
NNANETAOHPREMSMAAANAANAAYKAAAPVGLNLSALCCPGEHNAFRFP
FAGGSHHAAMAAVANGRPGCARPGGPGAGQNPRTISTORITTPAGGERIYPCVD
PYQGCPSAHTQMLKGAFCPCAGCTQCEHOKYGLAMASAGLPAPHPYQAANAANA
NAAARSAAPVYCSWVGADYCGKRRFQSDLEFSLRHTGNTSDPAAANAALAOSSOQ
SLTGTFPPSALIRAGYPTPLSPMSAANAARHYPAKPPGALAGSPFGAAGAFN
PAAANAALALGPYSPYAMYGRGCAAHQ"
59652..59723
/gene="BG:DS04641.2"
/note="symbol=BG:DS04641.2; evidence=predicted by content;
method=LRNscan-SE-1.0.1.0; method_score=57.22"
/product="tRNA-Pro"
/db_xref="FLYBASE:FBgn00"
59652..59723
/gene="BG:DS04641.2"
81865..81935
/gene="tRNA:G3:35Ba"
/note="symbol=tRNA:G3:35Ba; synonymy=BG:DS04641.3;
evidence=predicted by content; method=LRNscan-SE-1.0.1.0;
method_score=71.57"
/product="tRNA-Gly"
/db_xref="FLYBASE:FBgn0011862"
81865..81935
/gene="tRNA:G3:35Ba"
complement(82081..82151)
/gene="tRNA:G3:35Bb"
/note="symbol=tRNA:G3:35Bb; synonymy=BG:DS04641.4;
evidence=predicted by content; method=LRNscan-SE-1.0.1.0;
method_score=71.57"
/product="tRNA-Gly"
/db_xref="FLYBASE:FBgn0011863"

gene complement(82081..82151)
/gene="tRNA:G3:35Bb"
tRNA complement(82395..82465)
/gene="tRNA:G3:35Bc"
/note="symbol=tRNA:G3:35Bc; synonym=Bg:DS04641.5;
evidence=predicted by content; method=tRNAscan-SE-1.0;1.0;
method.score=71.57"
/product="tRNA-Gly"
/db_xref="FLYBASE:FBgn0011864"
gene complement(82395..82465)
/gene="tRNA:G3:35Bc"
complement(94320..94390)
/gene="tRNA:G3:35Bd"
/note="symbol=tRNA:G3:35Bd; synonym=Bg:DS04641.6;
Query Match 73.6%; Score 18.4; DB 6; Length 300962;
Best Local Similarity 95.0%; Pred. No. 3.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 6 tgcacgcacgattcgat 25
||||||| |||||||
Db 135310 tgcattgcctgcatttcgat 135329
RESULT 33
LOCUS D85415 9239 bp DNA BCT 07-MAR-2001
DEFINITION Pseudomonas putida gene for conversion of aniline to catechol.
ACCESSION D85415
VERSION D85415.1 GI:1841358
KEYWORDS
SOURCE Pseudomonas putida (strain:UCC22) DNA.
ORGANISM Pseudomonas putida
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (sites)
AUTHORS Fukumori,F. and Saint,C.P.
TITLE Nucleotide sequences and regulatory analysis of genes involved in
conversion of aniline to catechol in Pseudomonas putida
UCC22(PTDN1)
J. Bacteriol. 179 (2), 399-408 (1997)
JOURNAL
MEDLINE 97144524
REFERENCE 2 (bases 1 to 9239)
AUTHORS Fukumori,F. and Saint,C.P.
TITLE Aniline degradation in Pseudomonas putida UCC22(PTDN1): Initial
characterization of its conversion to catechol
Unpublished
JOURNAL
REFERENCE 3 (bases 1 to 9239)
AUTHORS Fukumori,F.
TITLE Direct Submission
JOURNAL
SUBMITTED (16-MAR-1996) Fumiyasu Fukumori, Toyo University, Faculty
of Life Sciences; 1-1-1 Izumino, Itakura, Ora, Gunma 374-0193,
Japan (E-mail:fukumori@itakura.toyo.ac.jp, Tel:0276-82-9212,
Fax:0276-82-9801)
FEATURES
source location/Qualifiers
1. 9239
/organism="Pseudomonas putida"
/strain="UCC22"
/db_xref="taxon:303"
1. 1703
/gene="tnpA"
<1..1703
/gene="tnpA"
/codon_start=3
/protein_id="BA12804.1"
/transl_table=11
/db_xref="GI:1841359"
translation="ALYIKGSRLKPEVAVASIGIMOVISSPDRERAPALVATL
PALRAVNGSWIEHLSIFGRARLFTDERWQESKRIARLSIPKATFTLPLL
ARVITGVDAVAAARSGLRVDDLEHLSPAEDEPVTGLRALDRIGEVOLPEV
ILADAVYFSNIMIGREPRSDDELAVYAGIMAGTSLTAVECARMIPOLSATSIRO
AMRWADRRRLSOACVALEFMORHPIATATWAGRSRLASSDMTMTETTKRVQARLDPR
RMTPSIGIYSHKRWGIFHAQPFVLRNAGVAIEGYIQEKLETSQLAVDTHTGYTD

gene 2802..4280
/gene="tdnQ"
CDS 2802..4280
/gene="tdnQ"
/note="amino group transfer"
/codon_start=1
/transl_table=11
/protein_id="BA12805.1"
/db_xref="GI:1841360"
translation="MSGKLEKRGIMSDPKAAADVLKIKERAGLOWRLSPDOYG
LHGKMLSVALLSAPASSSETMAPFFEDTSATYFNFPSADGIGSLAEGSRVY
MVPDPTFRLLPMDRTGMMADLWTSRPRPLRALILKALVEMDULGYDQGL
EYEWYLRIVDPCLBETLIGPCTPAAPKVPVAPVAGYSYLENHLDEVERIMAEVRQ
HLDALGMLPLRSIEDENWASQMETTFVPMFGLVDADTMVLEFRNAVQVCRRRGYLASF
KPAIIGPLASGMHLHQSILARDSGANAFIPQEGEALSALGSRVYGLLEHCAASF
TPTPTNGYRRRYSLSAPRVYAKNRAMA RYTSAPGDPASRVENRTGEPAAAPYL
YLASQVPSGIDGIRROLDPEPLOETPYAGDVTLLPNNSEALEVLETSKFFKEARGBE
FIRYMMHLRSEKRFVDAEGDYFSDPTVMEHREYELP"
4399..5160
/gene="tdnT"
CDS 4399..5160
/gene="tdnT"
/note="amino group transfer"
/codon_start=1
/transl_table=11
/protein_id="BA12806.1"
/db_xref="GI:1841361"
translation="MTAMKKYATWICSDASGDLLEOKMISAFRENEMETIOPAEN
NLEKRYDYSCHVITSGSPKSYDDAOTPLVSNLALRGAQGRBEPYGLCPGAOI
AAALGGEVGRNPSGRRKLGVDRLSENEAQLFPGVGAQPTVLVDSHCEVYTLIPG
GVOLASSORTIPHEVFLVNGQFLIGHPEDVROFLQOKPMAYHRAFLDDEWVQOE
SQODLDPREYIALGRLLDAGRILPATPODISALPA"
5195..6541
/gene="tdnA1"
CDS 5195..6541
/gene="tdnA1"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="large subunit of terminal dixygenase"
/protein_id="BA12807.1"
/db_xref="GI:1841362"
translation="MTTAKISVLPRTPAAGDMDDLVOEDRYHRLYTDEAIFSRM
NNIFATWVYLAHESLEPDEPKQAMIGREVIVARDEGVIRVSNRCSHRGASVC
REHNGNAGFTCPYHGWRPDRNGOLFGLGKNAYGTFPSRDMHLARPAOVASYKGF
PATLNDPAPSLAHLGNAPYLDANIDHNGSPENLPLAAGAPRLACNMKILVWDNAGD
GTHVPSHOSLWITNERVGGCMATFADDRSKRTNSALDNCHTYIDPREHGESA
WRLODPQGREPEYEDHVASLYGDAQOVLDTTVGAGMNLNIFENLALIGNOVAVDPL
AVASTGVHWATQKRDADPEVNTMRLTQEDPFVMEEMDAANFECQGLCNSPDE
WVDSRHHESGKVPVEDGIIIRGPTVTDLHMRYVQMKRLQAEPTLRMDGRILA"
6538..7182
/gene="tdnA2"
CDS 6538..7182
/gene="tdnA2"
/codon_start=1
/transl_table=11
/evidence=not experimental
/product="small subunit of terminal dixygenase"
/protein_id="BA12808.1"
/db_xref="GI:1841363"
translation="MSTATQAPAGRGYRHPQPSLYVVSFYDWLVDVYSADLARAV
REBQTVFAAREBDIVRLITVEARLLDQALTDQAYQVLALEECAYMTPAVSPAD
PQCSVTEPEFDRRLDRVTRGLTGALDSQFPTSPARQPSGLEWASGGRSEMAR
YSLTVLSRSGRGVLAGWNGFVLRTEAGSLIVLQVNLIDSDRQNNSPFL"
7194..8207
/gene="tdnB"
CDS 7194..8207
/gene="tdnB"
/db_xref="GI:1841364"
translation="MSTATQAPAGRGYRHPQPSLYVVSFYDWLVDVYSADLARAV
REBQTVFAAREBDIVRLITVEARLLDQALTDQAYQVLALEECAYMTPAVSPAD
PQCSVTEPEFDRRLDRVTRGLTGALDSQFPTSPARQPSGLEWASGGRSEMAR
YSLTVLSRSGRGVLAGWNGFVLRTEAGSLIVLQVNLIDSDRQNNSPFL"

```

/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="electron transfer protein"
/protein_id="BAI2809.1"
/db_xref="GI:1841364"
/translation="MAKRAOSLTITDVTAGSDAILSLRVDEQPKETPGQVILT
LAVVGQDEHMRCTSTSEVTPQPSIVLRVAGGSRVSMICDNAPRGLQVLP
GHFTLARGQPLLYAGSGSIAPFVALRALARARVRLFYARPRATAMLAELQ
ALQDAVGRLEIVMYDAEGLPTQAVLVAGQGLDQADAYMCGPPEPMHAGSLOV
AGFPAORVHREIDFCAVEGASEPAGGAPALITLYLMKGOTHAIPVAGELLISMLR
AGLPAHACRVBECASCMRLQAGEVORLDSVILDEDDVAAGMLACRTAASPALOY
RFS"
gene
8235..9142
/gene="tdnr"
8255..9142
/gene="tdnr"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="LysR-type regulator"
/protein_id="BAI2810.1"
/db_xref="GI:1841365"
/translation="MTRDPTPTPOMDLIRAFIALERHGSYVAELEGGIDSTLRRI
RLEOHGRALFVRAEGMRASADLNALISAORMEASPCODHAGACVIRISV
DVPRQRFAPFPAISEKYPKLVFTTTEAHFVMEQDVQVAVRLARPERNSALRV
KLGVAVGAYASDAVLERAAELAHENHQLAMNIQEPHODHPIYASLMSRGLSGO
VRQSDSFAPLAOLCALQGLALPKFLAAEYPOLVAHPNSVEVDITLWLVSRFDMA
AMORDLADMLQAEKARWPC"
9154..9239
/gene="tdnd1"
9154..9239
/gene="tdnd1"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="ferredoxin-like protein"
/protein_id="BAI19514.2"
/db_xref="GI:13364840"
/translation="MQATVAVLGPSSVCVWGTGTACSPGES"
BASE COUNT      1687 a      2889 c      2985 g      1678 t
ORIGIN

Query Match      72.8% Score 18.2; DB 2; Length 9239;
Best Local Similarity 87.0%; Pred. No. 5.4e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcagtgcatcgccacattcg 23
||||| ||||||| |||
Db 8837 GCAGTTCATCGCCACGACTTGG 8815

RESULT 34
AE004631 10864 bp DNA BCT 30-AUG-2000
LOCUS Pseudomonas aeruginosa PA01, section 192 of 529 of the complete
DEFINITION genome.
ACCESSION AE004631 AE004091
VERSION AE004631.1 GI:9948050
KEYWORDS Pseudomonas aeruginosa.
SOURCE Pseudomonas aeruginosa.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 10864)
Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S., Hutnagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Goltzy,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen

```

```

JOURNAL Nature 406 (6799), 959-964 (2000)
MEDLINE 20437337
REFERENCE 2 (bases 1 to 10864)
AUTHORS Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S., Hutnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E.,
Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
FEATURES
source
1..10864
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
/complement(111..1985)
/gene="PA2044"
/complement(111..1985)
/gene="PA2044"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAG05432.1"
/db_xref="GI:9948051"
/translation="MRRPSPFLAASVFCATPATALOPIAEAPLAGETPRCP
NADNSTDCVSTYSFTILKPSGRSMRSDRSVAFEDSLIVKKAELTQGGKPVLDLS
QIDRTAPNPOGFLERBOTSLAPNLAVGRISITTLREHTAPLSOPHYIISRP
MVRDRFVAERKAEPLFVSELMDAIRIQSDAKTKLVSKPOTTNTINAGNA
YLRTPRLIGSSLDLQNFGEFARVVEILAAELPGAAAANAAGKAPAEVAGL
MOYINDTVYRLGDMRSEGVYPSLAEIENGQDCKDLAILAAMIKAIGIAEPT
LYSRGVDVWDLVPGMYANHAIYVAEVDGKTMWIDPTNPVFAFRIMPDIQGMALY
LGADKVRDEIPLFAPGDTLVRVSEIYTHDGEARVESVELSNAPLMQSVADRQ
RGTSTDDLCENFKESGCVLEPHDDQFVLPSTYTSARITDRRALDRGCEVFR
ODLASQMPAFKRYRSEGLAELYICEPRTSYDLSLSGKTDPEAHSCETRSPPFDD
LOAEPAKDGGYHRYRYREVOKTSMLNHDEINAEFGKLIQSGVCEQLRLVKKDRP"
complement(2160..2420)
/gene="PA2045"
/complement(2160..2420)
/gene="PA2045"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAG05433.1"
/db_xref="GI:9948052"
/translation="MKRLIGLIRFYQYATSLIGPCRFYPSGSHYTLFAIRVHGL
RGYIGARRLLKCHPMHGGYDPPEROBQACACHTRAKPK"
complement(2479..2886)
/gene="PA2046"
/complement(2479..2886)
/gene="PA2046"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAG05434.1"
/db_xref="GI:9948053"
/translation="MGATRORLVELQEAADPRRLVIVQPRRTIGFAPAASEVRG
FYVLADGFFRQAYLGARNWVGETLIDVDVYLEGDAEAYCYGYSALIPRGAMREG
TLAPEPGLRGLATARSIVDGRITLYCAENAP"
complement(3921..4910)
/gene="PA2047"
/complement(3921..4910)
/gene="PA2047"
/codon_start=1
/transl_table=1
/product="probable transcriptional regulator"
/protein_id="AAG05435.1"
/db_xref="GI:9948054"
/translation="MPVAATRASLGKVSPPSPRSENTPLLSDLPALRROSELAQ

```

SLRHCPRDGIYGTAVEPLIALIRADGPTLPARGMHKPAALCIIAQGRKEVOIAEERYTY
 DPLHYVSYVTLPLAGOVITDSPPAPLVCVLDIDPAETIOLISDAPMEVASRRGR
 GLVYRIDASLDLVLRILHLDSPRODIPMLALIRLIFRLRSOGGOLHETALA
 DSOAHRTIRADIMNONGKPLRIEOLAVVNLSPTLHHFRKAVTAMSPLOYOKOLA
 LOEARRLIFSEGLEMAAGYVGESEPSQFSREYSRLGAPPLRLRARANTA
 5324..5731
 /gene="PA2048"
 5324..5731
 /gene="PA2048"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG05436.1"
 /db_xref="GI:9948055"
 /translation="MRHPYTPQRPAGVAPPALDLPVLAORRIHVEPASPMPNKA
 SHIVSIRARSQSOSHLRIQLQELAOAGAACGLRYELRQADGADLMLLHSEMSDEA
 AMOAYISGDAQRFVFAEVLMOALATLDOVERPF"
 complement(5789..7222)
 /gene="PA2049"
 complement(5789..7222)
 /gene="PA2049"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG05437.1"
 /db_xref="GI:9948056"
 /translation="MLVAIHGLIHRGLYGLHLCALVNDSEALYAHPATYRREDLL
 PAFESNEALREFEGVKKPLFTSANANHLVGLAEVLAAMLVWSPAPWLVRIEPP
 PGLDRAHARQGERRHTEAFRDGLPRVARALAHISLIMLITDDHDITDDNMISARWEL
 TAYGHPFSRRIIGNALVAYLLCOGNGNDPOHCEMLAVVAALMECAADGRLLPCAED
 RLIDALDPOHMYHVRPGSPPLIYDTRTRRMSRPNLGRPSGLMEALSEPODIL
 CEKATITSPAPMEGVKLEIGYQKLFSGISGIGOLYDANMAHRAASVMMNIRHSR
 TPGNLTIIISGDVHSFVYDIRIRHREGRPLMOLITSSIKNEPRRLIDMDRLMRKL
 YAPMSPLMWLQORLEIRPRVPOHSGOERLMSAGIGOVFIGEDGRPAVLQVNA
 GSPATAPFRHRVEDERADQSPVAPGR"
 8111..8617
 /gene="PA2050"
 8111..8617
 /gene="PA2050"
 /codon_start=1
 /transl_table=11
 /product="probable sigma-70 factor, ECF subfamily"
 /protein_id="AAG05438.1"
 /db_xref="GI:9948057"
 /translation="WMDOQTRNLVGLMPONDYSLSGRLRYLGGPSADIASETF
 AKVLAIPDPAISIREPALLTIRRLMYDGMWRDLEKAYLESILAGLEPALPSAEQ
 AGVETLLRLERMLAGLSPKARAFTHSQIGLIVYETACLELVSRSRIHOYMWGCFK
 OCYVLAE"
 8614..9567
 /gene="PA2051"
 8614..9567
 /gene="PA2051"
 /codon_start=1
 /transl_table=11
 /product="probable transmembrane sensor"
 /protein_id="AAG05439.1"
 /db_xref="GI:9948058"
 /translation="MSQGRERRILDEAAGMALLQSHAGAEERLAFOQWROADAENG
 AVEDRLIGOSFRNEPLHGLAEOLIGLEAPASRRRIQSLALAAVATLALGR
 IGIVGPMFGDLVTGTGROSRLPDEGSQLVIDASRVPEPDGRRRLRLREGCLLV
 DYARDPAPFVYDTGRGOVACQCTFOVSDGASRVVYLSRRVITTTADGRCCLER
 GGSARPDASAILARGPAGESAMLDGRLVENDRPLGEVLELRVRRKITSVADDA
 ALRVSGIYPLDSSVALALLESQSLPRLDYHGPVWSIASR"
 complement(9605..10075)
 /gene="cyns"
 /note="PA2052"
 complement(9605..10075)
 /gene="cyns"
 /codon_start=1
 /transl_table=11
 /product="cyanate lyase"
 /protein_id="AAG05440.1"
 /db_xref="GI:9948059"

gene
 CDS
 /translation="MOHSQVSPNARQOLAEFVUNKARLGLSMDLADGTGLATFEVT
 AALIGQHALPEANARARVAOLGDDDAVLLOSIPRGSIPGSDPTIYFEYPMLO
 VTSQILKALVHPEFGGIIISAINERKIDIKKVEDPEGGSNAVITTDGKVIYTPRP"
 complement(10115..10777)
 /gene="cynT"
 /note="PA2053"
 complement(10115..10777)
 /gene="cynT"
 /codon_start=1
 /transl_table=11
 /product="carbonate dehydratase"
 /protein_id="AAG05441.1"
 /db_xref="GI:9948060"
 /translation="MRDIIDGFLRFQORDAYPARSOLEKSLATROAPKALFIACSDSRV
 VPPELLQREGEELFVIRNMGNTIVYGPQPGVSAVEVAVANVIGDGVGSHSDG
 AMGALISQACLDLDPVAGLHLHAERANRANSHEHSAARLIDALVIRNVIYQIANTL
 RTHPCVARALEGRLNLHGWNVDIESGRIDALDGASRRVSLAHEPVAVGEPQGA
 VA"
 BASE COUNT 1603 a 3503 c 3970 g 1788 t
 ORIGIN
 Query Match 72.8%; Score 18.2; DB 1; Length 10864;
 Best Local Similarity 87.0%; Pred. No. 5.3e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 Oy 1 gaagtgatgcagcagattccg 23
 6964 GCAGGTACATCGCCAGCACCTCG 6986
 RESULT 35
 MTCY22D7/c 31859 bp DNA BCT 17-JUN-1998
 LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 133/162.
 DEFINITION Z83866 AL123456
 ACCESSION Z83866 AL123456
 VERSION Z83866.1 GI:3261691
 KEYWORDS
 SOURCE
 ORGANISM
 Mycobacterium tuberculosis.
 Mycobacterium tuberculosis
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacterineae; Mycobacteriaceae;
 Mycobacterium tuberculosis complex.
 1 (bases 1 to 31859)
 REFERENCE
 AUTHORS
 Cole, S.T., Brosch, R., Parkhill, J., Garrier, T., Churcher, C.,
 Harris, D., Gordon, S.V., Eigmeier, K., Gas, S., Barry III, C.E.,
 Tekala, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T.,
 Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S.,
 Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
 Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
 Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
 Squares, S., Squires, R., Sulston, J.E., Taylor, K., Whitehead, S. and
 Barrell, B.G.
 Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence
 Nature 393 (6685), 537-544 (1998)
 JOURNAL
 MEDLINE
 REMARK
 Erratum: [[published erratum appears in Nature 1998 Nov
 98295987
 12:396(6707):190]]
 2 (bases 1 to 31859)
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 DIRECT SUBMISSION
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced gi:1781132.
 NOTES
 Details of M. tuberculosis sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
 been renumbered from the original cosmid submissions but the old

gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gta, or tta) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

```

source
    1. .31859
    /organism="Mycobacterium tuberculosis"
    /strain="H37Rv"
    /db_xref="taxon:1773"
    <1. .29946
    /organism="Mycobacterium tuberculosis"
    /strain="H37Rv"
    /db_xref="taxon:1773"
    /clone="Y22D7"
    /complement(52. .504)
    /gene="nrdI"
    /complement(52. .504)
    /note="nrdI"
    /note="RV3052c, (MTCY22D7.30), len: 150. nrdI. Probable nrdI protein, similar to many e.g., sp1047415|NRDI_ECOLI NRDI PROTEIN (136 aa) fasta scores, opt: 574 z-score: 747.2 E(): 0; 62.2% identity in 135 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="nrdI"
    /protein_id="CAB06244.1"
    /db_xref="GI:1781162"
    /db_xref="SPTREMBL:P95107"
    /translation="MDIAGSLVSVSESVNTHREYOKIGTPATRIPLHGRIVDEPY VLIPTGGGRANGLDAGGYVPKOVATAPLNDHNRAGLRGVAAGTNGAFCYAG DVSRKCSVPYTRFELMGTEDDVAAARTGLAFWKQITGHQPSLQSL"
    /complement(518. .522)
    /note="possible RBS, AAGGA, for RV3052c"
    /complement(539. .778)
    /gene="nrdH"
    /complement(539. .778)
    /gene="nrdH"
    /note="RV3053c, (MTCY22D7.29), len: 79. nrdH. Probable nrdH protein, involved in electron transfer, shows strong similarity to other nrdH glutaredoxin-like proteins. FASTA best: sp1056108|NRDH_SALTY GLUTAREDOXIN-LIKE PROTEIN NRDH. (81 aa) fasta scores, opt: 243 z-score: 358.1 E(): 2e-12; 45.8% identity in 72 aa overlap"
    /codon_start=1
    /transl_table=11
    /product="nrdH"
    /protein_id="CAB06243.1"
    /db_xref="GI:1781161"
    /db_xref="SPTREMBL:P95106"
    /translation="MTVYVTRKPCVOCSTSKALDKGIAGVKVIDISLSEARDVYM ALGIOLPVPVAGNDHMSGFRPRIRKALAGALTA"
    /complement(789. .794)
    /note="possible RBS, AAGGAG, for RV3053c"
    /complement(1004. .1017)
    /note="conflict: this 14bp sequence is duplicated in BAC and w.g.s. data"
    /complement(1255. .1809)
    /gene="RV3054c"
    /complement(1255. .1809)
    /gene="RV3054c"
    /note="RV3054c, (MTCY22D7.28), len: 184. Function: unknown but shows some similarity to hypothetical bacterial proteins and contains aminotransferases class-II pyridoxal-phosphate attachment site (PS00599) near C-terminus. FASTA best: YIEF_ECOLI P31465 hypothetical 20.4 kd protein in tnab- (186aa) opt: 324, E(): 1.3e-14,
  
```

```

misc_feature
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein RV3054c"
    /protein_id="CAB06242.1"
    /db_xref="GI:1781160"
    /db_xref="SPTREMBL:P95105"
    /translation="MSDTRSDIKITLALGSLRAASFNRQIAFLAKKAPDGYTTFE GIADLPFTMEDID7ATVFPAPVSAUREASDAHAALVTPETNGSIPAVINADWLS RPPGDGALDKPLAYIGSGMGRYGVMAHDETRKFSIAGTRVDATIKLSVPPQTLGK SVADDGALANVDAVGNLAEEVG"
    /complement(1312. .1341)
    /gene="RV3054c"
    /note="PS00599 Aminotransferases class-II pyridoxal-phosphate attachment site"
    /complement(1818. .1822)
    /note="possible RBS, AAGGA, for RV3054c"
    /complement(1901. .2515)
    /gene="RV3055"
    /complement(1901. .2515)
    /note="RV3055, (MTCY22D7.26c), len: 204. Possible regulatory protein, low similarity to several. FASTA best: G987091 regulator of antibiotic transport complexes (204 aa) 90 opt: 150 z-score: 186.8 E(): 0.002, (28.9% identity in 197aa overlap). Contains potential helix-turn-helix motif from 497070(+3.29 SD)"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein RV3055"
    /protein_id="CAB06241.1"
    /db_xref="GI:1781158"
    /db_xref="SPTREMBL:P95103"
    /translation="MSGARIGDLPVFARQEPVPERGDAARNRALLLEARRLARSQ AATITMDVAAAGVCGKCTIPRRRSRAGIMVLLDEPERSCQAFLEGPPLPDAP PLDIRIATGREGMRVYHAAHOLLSANNDPOTRSHALSIVLRHLYLASAPTTGDL DQDTALLALLDVYEHQNLNGHFTLOTLDAMESLARKLCGR"
    /complement(2525. .3565)
    /gene="dinP"
    /complement(2525. .3565)
    /gene="dinP"
    /note="RV3056, (MTCY22D7.25c), len: 346. dinP, extensive similarity to proteins induced by DNA damage such as dinP, mcb, umuC. Also similar to MTCY48.28c. FASTA best: Y0JH_BACSU P54545 hypothetical 47.0 kd protein in glng- (414 aa) opt: 493, E(): 2e-26, (32.9% identity in 346 aa overlap), TParse score is 0.933"
    /codon_start=1
    /transl_table=11
    /product="dinP"
    /protein_id="CAB06240.1"
    /db_xref="GI:1781157"
    /db_xref="SPTREMBL:P95102"
    /translation="MPTAAPRMILHVDIOFLASVELLRHPELAGLPVIYGGNDPTE PRKVTGASVBARAVGAGMELTPAARCEAFELSNPAVAAAEVVALIRDQ YVPEWGDVDEAVLAAPCTPDDRIEVAIEIKVILSOTGSCISGIDNQRKATATG LAKPAGIQTLDANMAMTIGRTYBALMGVEPKTKRIAKIGITVQLAHTDSGLM STFGRTIALMLLAGGGDTEVSAQAWPERSRAVTFPRDLTCRSESAVVELAR TLNEVAVASSRTVRAVAVTRATETRTKIRKLOAPSTDDPVITAAARHVLDFEIDR PVRLIGVRLLELA"
    /complement(3619. .4482)
    /gene="RV3057c"
    /complement(3619. .4482)
    /gene="RV3057c"
    /note="RV3057c, (MTCY22D7.24), len: 287. Function: unknown but similar to various oxidoreductases and enzymes involved in polyketide synthesis. Contains short-chain alcohol dehydrogenase family signature (PS00061) and is related to other TB tuberculosis proteins e.g. MTCY190.25c, MTCY39.16c. FASTA best: Y00P_MCT0 Q10402 putative oxidoreductase cy190.25c (550 aa) opt: 462, E(): 3.9e-22; (34.5% identity in 258aa overlap). TParse score is 0.925"
  
```



```

/codon_start=1
/transl_table=11
/product="hypothetical protein RV3057c"
/protein_id="CAB06265.1"
/db_xref="gi:1781156"
/db_xref="SPTREMBL:P95101"
/transl_table="1"
/translation="MLORAGAGYFAGKRCFVWGAASGIGRATRLRLAOGAEILYDTR
DRDGLAQVCCNRALGAQVPEPRVYIDYDYDAAFADJHARHSMYVNLGVSA
WETVDQLTHDQSRWAIINLGPVHIVETLVPVPAAGRGHILNVSASAGLVGLPVR
AAVSASKYGLRGLSEVLRFDLARHGIGVSVVPGVAKPTLVNVEIAGVDDPRVNR
WVERSGHAYVPEKAAADKILAGVFRNRYLYVSADIRALYAFKRYAMWPTLYMRRVN
VFETRLRRGP"
complement(3949..4035)
/misc_feature
/gene="RV3057c"

Query Match      72.8%; Score 18.2; DB 3; Length 31859;
Best Local Similarity 87.0%; Pred. No. 5e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      2      caagtcacgcacgacattcga 24
|||||  |||||||  |||
Db      19871 CAGGTGCACGCCGACATCCGA 19849

RESULT 36
AC014955/c
LOCUS
DEFINITION
AC014955 103920 bp DNA HTG 16-NOV-1999
Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
AC014955
VERSION
AC014955.1 GI:6436380
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 103920)
REFERENCE
1 Adams,M. and Venter,J.C.
AUTHORS
Direct Submission
JOURNAL
Rockville, MD, USA
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
This sequence was identified as CDW:10212515 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..103920
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 29949 a 21507 c 21794 g 30670 t
ORIGIN

```

```

SOURCE
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 281869)
REFERENCE
1 Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amaratunga,C.P., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,B., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gaber
Man,K.H., Doyle,C., Abmayr,A., An,H.J.,
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktargolu,L., Beasley,E.M., Beeson,A.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Borchan,M.R., Bouck,J., Brokstein,P., Brotler,P., Burris,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Fowler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpis,G.H., Ke,Z.,
Kenshison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,J., Lei,Y., Levitsky,A.,
Li,J., Li,Z., Liang,Y., Lin,X., Lin,X., Liu,X., Matel,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Mishina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Murphy,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusker,D.R., Pacle,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Sanders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svrtkars,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
JOURNAL
20196006
TITLE
2 (bases 1 to 281869)
MEDLINE
REFERENCE
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.
AUTHORS
Direct Submission
JOURNAL
Rockville, MD, USA
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7294680.
FEATURES
source
1..281869
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3L"
/complement(5394..>6614)
/gene="CG12289"
/product="Ct19318"
/db_xref="FLYBASE:FBan0012289"
/db_xref="FLYBASE:FBgn0036160"
/db_xref="FLYBASE:FBgn0036160"
/complement(<5394..>6614)
/gene="CG12289"
/map="68C3-68C3"
/db_xref="FLYBASE:FBan0012289"
/db_xref="FLYBASE:FBgn0036160"
/complement(5394..6503)
/gene="CG12289"
/feature="CG12289 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012289"

```

```

/db_xref="FLYBASE:FBgn0036160"
/protein_id="AAF50065.1"
/db_xref="GI:7294728"
/translation="MAKAVLCGCTEVYVTTMDKVFQCDRCSTONCGMLORSGSSNV
STALRLGANVELFGVLTBNPTYRMILDDLEORGISDHSCTPDASPFSTIYODRPT
RCGVVYCDKSEFVYTAADPFKLDLDQGWNPFPARSPRETCMIRLLDHNNGRDR
DRIVYSLOIENSFEQVYVHILAMCYVYVTKYIARSCKMSFLEACBGNALNMPNRI
RIRRPCLIVPSSSLGACGMYTEGFEFVPAKPKIIDRVGSDSCFTAGIYAFVNO
RRLAEVDFANAVASYKLYGVFGDCLGHLIDAVKLPVRMRGSEVSDDDEMDQDLN
SCKHLLRPLEIFLEPKSIAPTKEVPIEFISLSS"
join(<6798..7180,7307..7723,7807..>7834)
/gene="CG7551"
/product="CT23115"
/db_xref="FLYBASE:FBan0007551"
/db_xref="FLYBASE:FBgn0036161"
<6798..>7834
/gene="CG7551"
/map="68C3-68C3"
/db_xref="FLYBASE:FBan0007551"
/db_xref="FLYBASE:FBgn0036161"
join(6798..7180,7307..7723,7807..7834)
/gene="CG7551"
/note="CG7551 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007551"
/db_xref="FLYBASE:FBgn0036161"
/protein_id="AAF50064.1"
/gene="GI:7294727"
/translation="MTQYGHKLMSAVGKKTVCILGTTTIDVSTIKRPPENSHQVVI
GGTWIRGKASNNCTVLANLGVKEFLGMLSSSVFQVAVDDLSKRGILINDCPICDO
GAPSSVILTKATKATRNIVYCNNNPEKEITVLSMEPDNNLDMMPLMDNDYVFEK
LADPNQWVSLDQCMQDERLRMRMGMLNLRKPYVYVLMGDOGAGIIDLNDNPFHVH
KPKRIVDALGAGDFVGAIFYALYIRERSVAVDGNRNASYKCTKNGQIKRML"
complement(join(<12006..12265,12321..13329,13402..>13545))
/gene="CG6140"
/product="CT19294"
/db_xref="FLYBASE:FBan0006140"
/db_xref="FLYBASE:FBgn0036162"
/evidence="not_experimental"
complement(<12006..>13545)
/gene="CG6140"
/map="68C3-68C3"
/db_xref="FLYBASE:FBan0006140"
/db_xref="FLYBASE:FBgn0036162"
/evidence="not_experimental"
complement(join(12006..12265,12321..13329,13402..13545))
/gene="CG6140"
/note="CG6140 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0006140"
/db_xref="FLYBASE:FBgn0036162"
/evidence="not_experimental"
/protein_id="AAF50063.1"
/db_xref="GI:7294726"
/translation="MSSTKKVETROESDPLGPMVEYMDRYGAGTMRCLINRIGEE
ERMPROIOMGLIKRAAETNOEBCIDRLKSLTASNAADDYISGLIYEGHPIDPIW
QTSQSTOSNNNSNEVIGNRAITELLSGRIGTKGDVHPNDVHNSQSSNDPFAIHIAV
ATALITDLRPAVTLRDSLOAKSEKWDIKIGRTHTODAVNLTLLQEFSGYAQOILIN
GLORIDAVLRVYQALAGTAVGTGINTRGFEKVKRILAOISGLFVAVANPEFAL
ACRDAMVEVHGLANLAVSLMKVTNDIRFGLSGPRGIGELFPENEPSSIMPQVN
PTQCEAMTNIACQVGNHVAIVSYGANGHELEINVEPLIASVATRSIKLLAGCSIFN
CNCVKGIKPKKEKLAKIVNESLMLVTALNPHIGYDKSQAIOAKAHRNGTTLLVEALNA
GISEKDFNEVNEREKMLGS"
join(<21646..22118,23232..>23712)
/gene="CG7547"
/product="CT23105"
/db_xref="FLYBASE:FBan0007547"
/db_xref="FLYBASE:FBgn0036163"
<21646..>23712
/gene="CG7547"
/map="68C3-68C4"
/db_xref="FLYBASE:FBan0007547"
/db_xref="FLYBASE:FBgn0036163"

```

```

CDS
join(21646..22118,23232..23712)
/gene="CG7547"
/note="CG7547 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007547"
/db_xref="FLYBASE:FBgn0036163"
/protein_id="AAF50062.1"
/db_xref="GI:7294725"
/translation="MATDAADKDRVASFEGSGMGAASATIFSAIGAAVYKSGGTGI
AAMAVRPLIKKSTIIPVMAGIITVIGLVSVIAGLSQSYTRKGYHILANGSV
GFAVAGPACITIGVDACVGRGAQOPRLFGVSHIILIFAEVGLGLVAILYTNKA
VTRALNEPSTVAFIAGCTGAAYAILFTTLGSGVTVAGSVGIKAMAVRPMIMKAIT
PVMKGTIATVGLVSVIAGLSIGDVTGDSYVHLAGLSVGLPGLTAGVAGIAGD
AGVKGTAQEPRLFVGMVILIFAEVLAIVGLVAILYTKL"
join(<26812..26991,28979..29261,30903..31335,32541..32677,
33043..33273,33411..33672,35647..35761,35930..>36106)
/gene="CG7543"
/product="CT23091"
/db_xref="FLYBASE:FBan0007543"
/db_xref="FLYBASE:FBgn0036164"
<26812..>36106
/gene="CG7543"
/map="68C3-68C4"
/db_xref="FLYBASE:FBan0007543"
/db_xref="FLYBASE:FBgn0036164"
join(26812..26991,28979..29261,30903..31335,32541..32677,
33043..33273,33411..33672,35647..35761,35930..36106)
/gene="CG7543"
/note="CG7543 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007543"
/db_xref="FLYBASE:FBgn0036164"
/protein_id="AAF50061.1"
/db_xref="GI:7294724"
/translation="MTPLDFYREOHLATKNHVEPNLSQVKTVAASQSRITRKSNSFA
NKTETMRROEGCGPHTVDPTQANRSQETPNPNVOLSSTIMPEKRRKSPLELEEK
RMKALNRFEEGRIDAELRPESEKRVGVGVVDPDTLLVGEYSGPSITIAOEL
PLPOAKOSINRLMIVVGGIRGAFGEFVGLVLAETAGVWATGPMACACRRHGV
LEELTHMALFMITSEFALQSKCKCKOCAPGKSSNPNPKOSRMRMSATKTRAGV
NEKALIRSTORSRRKRSKATYSRARVYLGGRORAVWDGIGMGODEVAVRANPVAND
AGQPPAPORLRLKWSIRRRRWCQMPWIDERIASLSNGSALVEASFQELRLEPFI
TFVLAKCICVSAVNWGFSRLATCGKSNHKTSQLNGAVLAAMVEVLTVALTG
GTADRLIOCSAKSLKRETTAKENEXTRRORSRRGLLEPPSLATPNETETETAE
RALSNEFVFRTRISEKRSRMAAKHSTSAHKONKN"
join(<36777..36839,42328..42970,43084..43194,43256..43751)
/gene="Charlyde"
/product="CT23077"

Query Match      72.8%  Score 18.2; DB 4; Length 281869;
Best Local Similarity 87.0%  Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      3 aggtgcattgccagcattcgat 25
Db      232720 ACGGCGTTTCGCCGACATTCGAT 232698

RESULT 38
LOCUS   MAV250020
DEFINITION MAV250020 2923 bp DNA BCT 26-SEP-2000
Mycobacterium avium subsp. paratuberculosis insertion sequence
15900, locus 7.
ACCESSION AJ250020
VERSION   AJ250020.1 GI:8919146
KEYWORDS  2599 gene; 2600 gene; insertion element; insertion sequence 15900;
p43 gene; transposase.
SOURCE   Mycobacterium avium subsp. paratuberculosis.
ORGANISM Mycobacterium avium subsp. paratuberculosis
Bacteria; Firmicutes; Actinobacterii; Actinobacteridae;
Actinomycetales; Corynebacteriinae; Mycobacteriaceae;
Mycobacterium; Mycobacterium avium complex (MAC).
REFERENCE 1 (bases 1 to 2923)

```

AUTHORS Bull T.J., Hermon-Taylor J., Pavlik I., El-Zaateri F. and Tizard M.
 TITLE Characterization of IS900 loci in mycobacterium avium subsp.
 JOURNAL Paratuberculosis and development of multiplex PCR typing
 MEDLINE Microbiology 146 (Pt 9), 2185-2197 (2000)
 REFERENCE 2 (bases 1 to 2923)
 AUTHORS Bull T.J.
 JOURNAL Direct Submission
 TITLE Submitted (29-NOV-1999) Bull T.J., Surgery, St. George's Hospital
 JOURNAL Medical School, London, SW17 0RE, UNITED KINGDOM
 FEATURES Location/Qualifiers
 source 1..2923
 /organism="Mycobacterium avium subsp. paratuberculosis"
 /sub_species="paratuberculosis"
 /db_xref="taxon:1770"
 /complement(11..466)
 /gene="2600"
 /complement(11..466)
 /gene="2600"
 /function="unknown"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="CAB96064.1"
 /db_xref="GI:8919147"
 /translation="MTAVEIGTYDLPVLRKAVATITLYFVCGRAVLVGFYAVDILT
 PGKRLQVFTDRPNVAVAGATATLVITATANSYSQGGGLGVAVVGMGVT
 LLGVALLAMHLLPGSFHEVEEPQLHPSFAVALLLAVGVTAAVS"
 473..1922
 /organism="Mycobacterium avium subsp. paratuberculosis"
 /insertion_seq="IS900"
 /sub_species="paratuberculosis"
 /db_xref="taxon:1770"
 /gene="p43"
 /complement(1906..1906)
 /gene="p43"
 /codon_start=1
 /transl_table=1
 /product="transposase"
 /protein_id="CAB96065.1"
 /db_xref="GI:8919148"
 /translation="MTVEYVVAQPVWAGVADGKADHYCWINDDAORLLSORVANE
 AALLEIAAVTTADGGEVFWAIDNAGALLALLIAGORLLYIPGRVHAAS
 YRGSKTDADKADITADARRHDLPLRACDDIIVELRITTSRSLVADRTRAIIP
 NARPAAGILSLERAFYKNSRAALILITGTDPDLRSAGAVAAIIEKRRKRRND
 VAAATALQANNAHSIVGQDLATVVARLKEVMALDTEIGDDAMIIEKRRKRRND
 EILSMGFGVILCAEFLAATGDMAFASDRLAGVAGLAPVRDGRISGNLRRP
 RYDRRLRACVLSLIRTDPSSTRYYDRKRTEGKHTQAVIALARRRLNLVAMLR
 DHAYYHPTTAA"
 complement(1963..2394)
 /gene="2599"
 /complement(1963..2394)
 /gene="2599"
 /function="unknown"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="CAB96066.1"
 /db_xref="GI:8919149"
 /translation="VSRNRLFLVIGALAAVAVCVLIGIILLQKNIASYVAGHYEVA
 RDVNGTRQCTGSSPOVADTLSEVAEPARADDDKTEYLRSNDIVIVGPDGNVPCSI
 RVEPLSNXYSIRAEVFLGPGFPGSPSGAGCTGGCGTK"
 BASE COUNT 473 a 1005 c 1000 g 445 t
 ORIGIN

Query Match 72.0%; Score 18; DB 3; Length 2923;
 Best Local Similarity 100.0%; Pred. No. 71e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 gcaggtcaccgacagca 18
 |||

Db 127 GCAGGTCCATCCGACGACA 144
 RESULT 39
 G01446
 LOCUS
 DEFINITION Dm0466 Drosophila p1 library Drosophila melanogaster STS genomic
 clone DS07967 SP6, sequence tagged site.
 ACCESSION G01446
 VERSION G01446.1 GI:684849
 KEYWORDS
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 RUBIN, G.
 1 (bases 1 to 479)
 REFERENCE
 AUTHORS Rubin, G.
 TITLE Berkeley Drosophila Genome Project
 JOURNAL unpublished (1994)
 COMMENT
 Contact: Berkeley Drosophila Genome Project
 Primer A: CTCCTCCTGCGACTTCCTA
 Primer B: CACCAAGGCGCATATCCGA
 STS size: 192
 PCR Profile:
 Annealing: 58 degrees C PCR Cycles: 30
 Protocol:
 Template: p1 library pools
 Primer: 1 uM each
 dNTPs: 250 uM each
 Taq Poly: 0.05 units/u1
 Total Vol: 15 u1
 Buffer:
 MgCl2: 1.5mM
 KCl: 50 mM
 Tris-HCl: 50 mM
 pH: 8.3
 Gelatin: .001 %
 The p1 library has been distributed to 16 regional sites. A list
 of these sites is available from FlyBase, via anonymous ftp to
 ftp.bio.indiana.edu in the file
 flybase/allied-data/genome-projects/1b1/BLB.doc.
 Location/Qualifiers
 1..479
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="DS07967"
 /clone_lib="Drosophila p1 library"
 /note="Vector: PAD10Sacr1. The p1 library was made by D.
 Smoller in D. Hartl's lab (see Smoller et al., Chromosoma
 100: 487). Chromosomal position was mapped by polytene
 chromosome in situ hybridization in the Hartl lab (see
 Hartl et al., PNAS 91: 6824). STS's were generated by
 sequencing the ends of the Drosophila insert in these p1
 clones, by the W. Kimmerly, C. Martin, and M. Palazzolo
 lab at LBL."
 1..250
 STS
 primer_bind complement(233..250)
 BASE COUNT 113 a 120 c 97 g 145 t 4 others
 ORIGIN

Query Match 71.2%; Score 17.8; DB 54; Length 479;
 Best Local Similarity 90.5%; Pred. No. 9.9e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 3 aggtgcatcgccagcattcg 23
 |||
 Db 392 ATGGCATTGCGCAGCATTTGC 412

```

RESULT 40
MSU31280      3748 bp    DNA          BCT      21-NOV-1996
LOCUS        Methylobacillus sp. beta-lactamase (bla) and biotin synthase (biob)
DEFINITION   genes, complete cds.
ACCESSION    U31280
VERSION      U31280.1  GI:1680651
KEYWORDS
SOURCE       Methylobacillus sp. KTI.
ORGANISM     Methylobacillus sp. KTI
Bacteria: Proteobacteria; beta subdivision: Methylophilus group;
Methylobacillus.
REFERENCE    1 (bases 1 to 3748)
AUTHORS      Serebrijski,I., Vassin,V.M. and Tsygankov,Y.D.
TITLE        Two new members of the bio B superfamily: cloning, sequencing and
              expression of bio B genes of Methylobacillus flagellatum and
              Corynebacterium glutamicum
JOURNAL      Gene 175 (1-2), 15-22 (1996)
MEDLINE      97074643
REFERENCE    2 (bases 1 to 3748)
AUTHORS      Serebrijski,I., Vassin,V. and Tsygankov,Y.
TITLE        Direct Submission
JOURNAL      Submitted (10-JUL-1995) Ilya Serebrijski, Institute of Genetics and
              Selection of Industrial Microorganisms, Ist Dorozhny Proezd,
              Moscow, 113545, Russia
              Location/Qualifiers
FEATURES
Source       1..3748
              /organism="Methylobacillus sp. KTI"
              /isolate="KTI"
              /db_xref="taxon:79675"
              201..1061
              /gene="bla"
              201..1061
              /gene="bla"
              /codon_start=1
              /transl_table=11
              /product="beta-lactamase"
              /protein_id="AAC44581.1"
              /db_xref="GI:1680652"
              /translation="MSIOHFRVALIPFEAFCLPVPAPHETLVKVKADAEQDLAGRVY
              IELDLSNGKILIESFPERPEPMSTFKVLGAVLSRIDGOGOLGRRIHYSNDLVE
              YSPVTEKHLDGAVRELCSAITMSDNTAANLLTTIGGPKELTAFLHMGCHYRL
              DRWPELNEAIPNDEROTTPVAMATIRKLITGELLTASRQOLDIMAEADVACPL
              LRSALPWGFIADKSGAGERSGIIAALCPDCKPSRIYVITTTGSQATMDERNROIA
              EIGASLIRHW"
              2141..2153
              /gene="biob"
              2141..3243
              /gene="biob"
              2177..2183
              /gene="biob"
              2207..2211
              /gene="biob"
              2218..3243
              /gene="biob"
              /note="Method: conceptual translation supplied by author"
              /codon_start=1
              /transl_table=11
              /function="sulfur inserting enzyme; last step of biotin
              biosynthesis"
              /product="biotin synthetase"
              /protein_id="AAC44582.1"
              /db_xref="GI:1680653"
              /translation="MTMTNSHRMDSVIDINGLDNRVQARVQAEAPKRWSDVDIV
              ALPELPSDLMRADSVARENFPNGVOVSTLSTKTGCSGDCGCPQAAAYHTDVE
              KODLHAAGRTSRVLPFRMRQPCMAAWQPKRDLEPLVAMIREVKAMGETCATLG
              MKDGAQDLKESAGLDYNNHNDTAPPEYGEVITTRTYODRLDTRVREDDINVCCG
              GIIGMGSRVORAGLILANMERPEVDPDLTVGEGTPMGMDPLDPEFEVTTA
              AARTTPOSFVRLSVGRQSMHSGIALCTLAAGANSLFTGEEKLITTGNEAEADKRLFD
              KIGIHPL"
BASE COUNT   895 a      980 c      988 g      885 t

```

```

ORIGIN
Query Match      71.2%; Score 17.8; DB 3; Length 3748;
Best Local Similarity 90.5%; Pred. No. 8.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 5 gtagcgcgcagcattcgat 25
|||||||  || |||||
Db 2398 GTGCATCGCGAGATTTTCGAT 2418

RESULT 41
CVU39574
LOCUS        CVU39574      3748 bp    DNA          circular SYN      25-NOV-1995
DEFINITION   Expression cloning vector pBV102, including biotin synthetase
              (biob) and beta-lactamase (bla) genes, complete sequence.
ACCESSION    U39574
VERSION      U39574.1  GI:1073246
KEYWORDS
SOURCE       Cloning vector pBV102.
ORGANISM     Cloning vector pBV102
              artificial sequence; vectors.
REFERENCE    1 (bases 1 to 3748)
AUTHORS      Serebrijski,I., Vassin,V. and Tsygankov,Y.
TITLE        Cloning, sequencing and expression of bioB genes of Methylobacillus
              flagellatum and Corynebacterium glutamicum
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 3748)
AUTHORS      Serebrijski,I., Vassin,V. and Tsygankov,Y.
TITLE        Direct Submission
JOURNAL      Submitted (27-OCT-1995) Vitaly Vassin, Institute of Genetics and
              Selection of Industrial Microorganisms, Ist Dorozhny Proezd, 1,
              Moscow, 113545, Russia
              (U31280).
COMMENT
FEATURES
Source       1..3748
              /organism="Cloning vector pBV102"
              /db_xref="taxon:43991"
              /note="biob gene expression vector based on pUCBM20
              replicon; biob gene was placed under the control of the
              lac promoter; production of bioB protein in
              E.coli[pBV102] cells reaches up to 25% of the cell
              protein"
              201..1061
              /gene="bla"
              201..1061
              /gene="bla"
              /codon_start=1
              /transl_table=11
              /product="beta-lactamase"
              /protein_id="AA81772.1"
              /db_xref="GI:1073247"
              /translation="MSIOHFRVALIPFEAFCLPVPAPHETLVKVKADAEQDLAGRVY
              IELDLSNGKILIESFPERPEPMSTFKVLGAVLSRIDGOGOLGRRIHYSNDLVE
              YSPVTEKHLDGAVRELCSAITMSDNTAANLLTTIGGPKELTAFLHMGCHYRL
              DRWPELNEAIPNDEROTTPVAMATIRKLITGELLTASRQOLDIMAEADVACPL
              LRSALPWGFIADKSGAGERSGIIAALCPDCKPSRIYVITTTGSQATMDERNROIA
              EIGASLIRHW"
              2130..2186
              /note="lac promoter"
              2207..3243
              /gene="biob"
              2207..2211
              /gene="biob"
              2218..3243
              /gene="biob"
              /codon_start=1
              /transl_table=11
              /product="biotin synthetase"
              /protein_id="AA81773.1"
              /db_xref="GI:1073248"
              /translation="MTMTNSHRMDSVIDINGLDNRVQARVQAEAPKRWSDVDIV

```

ALFELPSDLMHRAOSVHRENDPENGVOVSTLSTIKGSCSEDOGYOPARHYTIVE
KODHAGRCGRSVLIPRMROPCMAAMOPKORDLPPVLAIRYKAMLETCATLG
MLKDGAEQJKEAGLDYNNNDTAPRYGVEITTRTYQDRUTLDRVRQDINVCQG
GIIOMGESRVORAGLLAQLANMERPEPSVDPOLITVOEGPMPGMDLDPFEVRIIA
AARTMPOSFVRSLVSGROSMEGICQALCFLAGANSIFYGKELITGTGNEAEADRKLFD
KLGIHPL"

misc_feature 2245..3315
/note="Isolated from *Methylobacillus flagellatum*"
misc_feature join(3321..3748,1..2232)
/note="derived from pUC18"

BASE COUNT 895 a 980 c 988 g 885 t
ORIGIN

Query Match 71.2%; Score 17.8; DB 56; Length 3748;
Best Local Similarity 90.5%; Pred. No. 8.7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 gtgcacgcacgattcgat 25
|||||
Db 2398 GTGATCGGAGATTTCAT 2418

RESULT 42
CELFA1H10 35928 bp DNA INV 06-APR-2001
LOCUS
DEFINITION *Caenorhabditis elegans* cosmid F41H10, complete sequence.
ACCESSION U61954
VERSION U61954.2 GI:4883513
KEYWORDS HTG.
SOURCE *Caenorhabditis elegans*.
ORGANISM *Caenorhabditis elegans*.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilita;
Rhabdilitida; Rhabdilitae; Peleiderinae; *Caenorhabditis*.
1 (bases 1 to 35928)
The *C. elegans* Genome Sequencing Consortium, Washington University
Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,
Hinxton, U.K.C.
TITLE Genome sequence of the nematode *C. elegans*: a platform for
investigating biology. The *C. elegans* Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REFERENCE 2 (bases 1 to 35928)
AUTHORS Blanchard, M. and Bradshaw, H.
TITLE The sequence of *C. elegans* cosmid F41H10
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 35928)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-1996)
REFERENCE 4 (bases 1 to 35928)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 35928)
REFERENCE Waterston, R.
AUTHORS Direct Submission
TITLE Submitted (06-APR-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 21, 1999 this sequence version replaced gi:1397328.
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RO, England
e-mail: tw@emmatode.wustl.edu and jesus@sanger.ac.uk

neighboring submissions.

WARNING: These data have only had automated annotation and have not yet been subjected to manual review of that annotation. We will be manually reviewing this information as quickly as possible and at that time this GenBank record will be updated and this warning removed.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES

source

gene
CDS
complement(100..291,623..910,1364..1645,2158..2520,2580..2714,5664..6123,6419..6973,7024..7292,7342..7629,7674..7709)
/gene="F41H10.6b"
/note="coded for by *C. elegans* cDNA yk3c3.3; coded for by *C. elegans* cDNA yk4c9.3; coded for by *C. elegans* cDNA yk15a12.3; coded for by *C. elegans* cDNA yk15a12.5; coded for by *C. elegans* cDNA yk23b11.3; coded for by *C. elegans* cDNA yk23b11.5; coded for by *C. elegans* cDNA yk44h4.3; coded for by *C. elegans* cDNA yk44h4.5; coded for by *C. elegans* cDNA CEMSE39f"
/codon_start=1
/product="Hypothetical protein F41H10.6b"
/protein_id="AAK29810.1"
/db_xref="GI:13559598"
/translation="MLPDRKRNKSTGPIPLIFPNOTONENHTVCPTRPESDRILKI
KALTKTKILECTVLTNLEIDADLETHHNSWVKDMESEKTOEDINSQCKYD
SYFEMENSMKAVKAGVACVDRDLNRMAMEASGFAVVRPGHNSVSPGCFLENN
VQAAEAEPFSAERLLIYDLVNHGHGFORLEFDKRYLPSIRHNEHGLPMLPE
SDPDHIGSGKGLYNNALNLNMGCTSDSYLSIFHVLPLATOPRPHRYIASGFA
LGDPLGKCLIPDGSHTLYHLKSLAGSRMLVVEGCGNHQISAAVQRCVRLCY
APPSLELNEAPRSTYDSCVLSVLRHMCNCFDPPSTSLRLQMPVTKVITYN
DPTTRRADTGEIIOBELASTFTASDVLPETMMETLIYFNESDDAHPLEDNHPEK
ARTRIILKTLIRBSGVLEKCVDRNCERATNEIRLVIHKMLEHRLRTETMDEIME
EAKKEFNSIYLRDLTKAVAKAVAGVLOSDEIFKEDAGRNALYVRPGHNASAK
SSGFCIFNNVAAAKYAOBRHAKRYLLIDMDVNHGNGOELFEYDSNMVMSIHRD
KGNFYPIGPKYSDYEGAGSGMSYNNPFSVQMGDNFYQMAFPQRYMPLAYQRPND
LVILSAGFPAADVDPLGETKVPETRALMTVQLSSLAGRIITVLEGGINTLSISNA
QAVCEVLQNRSMRLRLREKEQFATKPKIETSSCIKTITREYCAVQOKWLSILKGQV
PNSYGLIDIDEAVYDSDIDMADSSSSSSSTPRSHLEIMDSGPAHAAVPLATCP
HIKEVYLPKAPNARTACSECOIGAEVWTCLYCKYNGRGYVNEHAMHILSSHPM
ALSMADLSWYCPDSTVYNNPALIGAKSAHNSKGTETPS"
complement(1094..7709)
/gene="F41H10.6a"
complement(join(1094..1297,1364..1645,2158..2520,2580..2714,5664..6123,6419..6973,7024..7292,7342..7629,7674..7709))
/gene="F41H10.6a"
/note="coded for by *C. elegans* cDNA yk1c4.5; coded for by *C. elegans* cDNA yk270d8.3; coded for by *C. elegans* cDNA yk270d8.5"
/codon_start=1
/product="Hypothetical protein F41H10.6a"
/protein_id="AAK29809.1"
/db_xref="GI:13559597"
/translation="MLPDRKRNKSTGPIPLIFPNOTONENHTVCPTRPESDRILKI
KEALTKTKILECTVLTNLEIDADLETHHNSWVKDMESEKTOEDINSQCKYD
SYFEMENSMKAVKAGVACVDRDLNRMAMEASGFAVVRPGHNSVSPGCFLENN
VQAAEAEPFSAERLLIYDLVNHGHGFORLEFDKRYLPSIRHNEHGLPMLPE
SDPDHIGSGKGLYNNALNLNMGCTSDSYLSIFHVLPLATOPRPHRYIASGFA

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between

```

LLGDLPGMCLTPDGYSHILYHLKSLAOGMLVYLBEGVNHQISAVAVQRCVLLCY
APFSLLENAPKESTVDSVSVLSVLHNNMCPEYPSRSTSLRAQMPVNTVITNY
DPTTRRADTGEI10DELASTETASDIPTENMETLTYFNEGDAHDELDHEMKEP
ARRIRIKTLRESGVELEKVCVRCNCRERATNEIELVHTKMLERHTTEEMKEDELE
EAKRENSIYLTREDTLVKARAVAGAVLOSDEIEKADAGORNALVIRPBGHNASAK
SSGCFITNNVAVAKVQORHAKARVILDMVDHNGTQETIEFEDNSVWMSIHRD
KGNFYPIGEPRKSDVCEGACGECMSVVPSCVOMGNEVOMAFORITMIAVQPNF
LVLSAGDAVADDPDCEKYVTEFTALMTYOUSLHGRIYVLBGVNLSTISNA
OAVCEVLQNRSMLEKREK EOPATKPOKTESCIKTRREVCAVOOKYWSILKGFOII
IGSVLNSKRLINKNKSAAVLKVGKATDPVKAODSDRSRYNTRRRRSANDEVDWE
KLENNMKT"
/gene="F41H10.5"
4501..5564
join(4501..4670,4739..5062,5225..5564)
/gene="F41H10.5"
/codon_start=1
/evidence="not-experimental"
/product="Hypothetical protein F41H10.5"
/db_xref="GI:13559592"
/translation="MSEETSSLSLRIRYRPVNLNFIPLNIPLVYSLPFRMAY
VKANFWLKEFGKRYRLQNLVQWLAHVIVGGDLILMFSPHEIPHDVIDWNPILAA
QLPLISVAYFEKADFMALMYEMNSYTELLELHATCAALVAGLGGFLAGMALI
MEVNSILHATIIQICLSTOEPGVVAVYSLIIEFFRISQILMTNMAIYVNP
GLHWYVYVGGPIIFGCIQMLFLRLASDGLFELKARFAMTRDKNDDEKKKG
N"
8048..10068
/gene="F41H10.4"
join(8048..8221,8267..8392,8440..8587,8640..9481,
9530..9643,9696..9858,9908..10068)
/gene="F41H10.4"
/note="short region of weak similarity to Plasmodium
yoelii rhoptry protein (PIR:C45521); coded for by C.
elegans CDNA CE5B49F; coded for by C. elegans CDNA
CE5K03F; coded for by C. elegans CDNA yk707f2.3"
/codon_start=1
/protein_id="AAK29812.1"
/db_xref="GI:13559600"
/translation="MATESGGLSKEEFLVQPOLLOKNEVLEIREIKKMAASOH
ONSPKNALQFASKLINRASSKKDEAEITSLRKLTDQOEIRLOOSTIFELKITS
TONEITLKNQSSYSDHPLEKAEERVOEDLKEKIKITDEKVLCOETIPALSNKY
SLSELENCQTLNVEIDRLSLITSESEALIKIEKVEENLRIKIEIELRNGEI
AEKKESNNVAVMDEKKLSDNLEKRTMLQSKSEIDSLKAENFKLASEQLNHDQSK
NLAIQSEKLTLEISQOETISGLKRIEVLKSNHRSIQOONTNAAHNLKLEED
RVLLERVOLEVDLEKSPESDECELTMTSLTEELKSAEDSKLAVALKASLVE
LQSLKEERKADSYERKSEEPAGWYHPRENDROSHPDGMSVSKATSEAVE
LIRPLATLOKTHSENADSLIQLESNRLKRYTEKGELENIILREKPICTGQOON
LGSSSSSVQPKLEASFRKLNLTLGDSRSDVRENNKRLQRLMEETLSKNILLODL
QTIEMERTGL"
10354..10355
/note="SL2 trans-splice site; see ESTa yk746c5.5 and
yk733e6.5"
10356..12210
/gene="F41H10.11"
join(10356..10546,10605..10759,10854..10913,11076..11428,
11479..11885,11937..12210)
/gene="F41H10.11"
/note="coded for by C. elegans CDNA cm12c12; coded for by
C. elegans CDNA yk746c5.5; coded for by C. elegans CDNA
yk733e6.5"
/codon_start=1
/product="Hypothetical protein F41H10.11"
/protein_id="AAK29811.1"
/db_xref="GI:13559599"
/translation="MEVEVPGSGSEADTGSSEASYKSOVEIEDEENLKNADIFEHL
EOLPEOFTLSEFGKPIFVANNDRNGEIVLALVAFAPRSQMDSDIMTMSQNH
IOELHKSPLIFCVVRADLANCODEPFFVAFKFSKYPRDQDLQVLYRQISLISK
SOLENYKRRKGGNDLRLKLRIGDRLLDSTSSWRSPINLVDSISALPWNDSDEF
LSTTMSCLGAAKIDGALFGLIARQIATAVAFKKYMHPRDLNVLVLSQNTLOT
DSQWVPICLPRENDGTFYAYISYPCNKEDQIPQIVLISYKRHPGLKLVKROQI
VTLENNQKFFVNAQAMKTPNLQIOLISQNSGSLFVLHSSKQVMSKSLPL
ITREERWIASEMRSSTSLHPLRTTLEFVCRTKCLFLWVTDLSLVCIFGPFVTA
TIAVOVERKLKLSKSHQRYFTINSFS"

```

```

gene 13294..17897
CDS /gene="F41H10.3"
join(13294..13479,13539..13592,13639..13719,13769..13918,
13972..14274,14324..14462,14885..15450,15512..16055,
16174..16650,16701..16979,17155..17635,17735..17897)
/gene="F41H10.3"

Query Match 71.2%; Score 17.8; DB 6; Length 35928;
Best Local Similarity 90.5%; Pred. No. 7.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ggtgcacgcacacattcga 24
|||||

Db 34655 GTTGCAATCCACGACATTTCGA 34675

RESULT 43
AC019940/c 53067 bp DNA HTG 03-JAN-2000
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, In ordered
DEFINITION
AC019940
AC019940
AC019940.1 GI:6664957
HTG: HTGS_PHASE2.
KEYWORDS
fruit fly.
SOURCE
Drosophila melanogaster
ORGANISM
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 53067)
AUTHORS
Adams,M. and Venter,J.C.
JOURNAL
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
This sequence was identified as CDM:10211463 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..53067
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 14944 a 11879 c 11720 g 14524 t
ORIGIN

Query Match 71.2%; Score 17.8; DB 65; Length 53067;
Best Local Similarity 90.5%; Pred. No. 7.4e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 agtgcacgcacacattcga 23
|||||

Db 13633 ATGCGCATGCCACGACATTTCG 13613

RESULT 44
AC016014/c 68419 bp DNA HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-26J10, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
AC016014
AC016014
AC016014.2 GI:9134660
HTG: HTGS_PHASE0.
KEYWORDS
human.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 68419)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL
Homo sapiens, clone Rp11-26J10
Unpublished
2 (bases 1 to 68419)
REFERENCE

```

AUTHORS

Birtten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karas,A., Klein,J.,
 Lehoczek,J., Liu,C., Locke,K., MacDonald,P., Marquis,N.,
 McGraw,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Tsefaye,S., Tittell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE
JOURNAL

COMMENT

Submitted (18-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 13, 2000 this sequence version replaced gi:6449522.
 All repeats were identified using RepeatMasker:
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Center project name: L4643

Center clone name: 26_J_10

NOTE: This record contains 76 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 788: contig of 788 bp in length
 * 889 888: gap of 100 bp
 * 1667 1766: contig of 778 bp in length
 * 1767 2567: contig of 801 bp in length
 * 2568 2667: gap of 100 bp
 * 2668 3636: contig of 969 bp in length
 * 3637 3736: gap of 100 bp
 * 3737 4595: contig of 859 bp in length
 * 4596 4695: gap of 100 bp
 * 4696 5517: contig of 822 bp in length
 * 5518 5617: gap of 100 bp
 * 5618 6436: contig of 819 bp in length
 * 6437 6536: gap of 100 bp
 * 6537 7361: contig of 825 bp in length
 * 7362 7461: gap of 100 bp
 * 7462 8283: contig of 822 bp in length
 * 8284 8383: gap of 100 bp
 * 8384 9171: contig of 788 bp in length
 * 9172 9271: gap of 100 bp
 * 9272 10074: contig of 803 bp in length
 * 10075 10174: gap of 100 bp
 * 10175 10962: contig of 786 bp in length
 * 10963 11062: gap of 100 bp
 * 11063 11845: contig of 783 bp in length
 * 11846 11945: gap of 100 bp
 * 11946 12758: contig of 813 bp in length
 * 12759 12858: gap of 100 bp
 * 12859 13619: contig of 761 bp in length
 * 13620 13719: gap of 100 bp
 * 13720 14542: contig of 823 bp in length
 * 14543 14642: gap of 100 bp

14643 15428: contig of 786 bp in length
 * 15429 15528: gap of 100 bp
 * 15529 16328: contig of 800 bp in length
 * 16329 16428: gap of 100 bp
 * 16429 17221: contig of 793 bp in length
 * 17222 17321: gap of 100 bp
 * 17322 18099: contig of 778 bp in length
 * 18100 18199: gap of 100 bp
 * 18200 18960: contig of 781 bp in length
 * 18961 19080: gap of 100 bp
 * 19081 19869: contig of 789 bp in length
 * 19870 19969: gap of 100 bp
 * 19970 20807: contig of 838 bp in length
 * 20808 20907: gap of 100 bp
 * 20908 21769: contig of 862 bp in length
 * 21770 21869: gap of 100 bp
 * 21870 22649: contig of 780 bp in length
 * 22650 22749: gap of 100 bp
 * 22750 23545: contig of 796 bp in length
 * 23546 23645: gap of 100 bp
 * 23646 24442: contig of 797 bp in length
 * 24443 24542: gap of 100 bp
 * 24543 25345: contig of 803 bp in length
 * 25346 25445: gap of 100 bp
 * 25446 26227: contig of 782 bp in length
 * 26228 26327: gap of 100 bp
 * 26328 27126: contig of 799 bp in length
 * 27127 27226: gap of 100 bp
 * 27227 28039: contig of 813 bp in length
 * 28040 28139: gap of 100 bp
 * 28140 28965: contig of 826 bp in length
 * 28966 29065: gap of 100 bp
 * 29066 29854: contig of 789 bp in length
 * 29855 29954: gap of 100 bp
 * 29955 30736: contig of 782 bp in length
 * 30737 30836: gap of 100 bp
 * 30837 31593: contig of 757 bp in length
 * 31594 31693: gap of 100 bp
 * 31694 32483: contig of 790 bp in length
 * 32484 32583: gap of 100 bp
 * 32584 33391: contig of 808 bp in length
 * 33392 33491: gap of 100 bp
 * 33492 34285: contig of 794 bp in length
 * 34286 34385: gap of 100 bp
 * 34386 35223: contig of 838 bp in length
 * 35224 35323: gap of 100 bp
 * 35324 36128: contig of 805 bp in length
 * 36129 36228: gap of 100 bp
 * 36229 37039: contig of 811 bp in length
 * 37040 37139: gap of 100 bp
 * 37140 37899: contig of 760 bp in length
 * 37900 37999: gap of 100 bp
 * 38000 38778: contig of 780 bp in length
 * 38780 38879: gap of 100 bp
 * 38880 39673: contig of 794 bp in length
 * 39674 39773: gap of 100 bp
 * 39774 40594: contig of 821 bp in length
 * 40595 40694: gap of 100 bp
 * 40695 41475: contig of 781 bp in length
 * 41476 41575: gap of 100 bp
 * 41576 42363: contig of 788 bp in length
 * 42364 42463: gap of 100 bp
 * 42464 43252: contig of 789 bp in length
 * 43253 43352: gap of 100 bp
 * 43353 44175: contig of 827 bp in length
 * 44180 44279: gap of 100 bp
 * 44280 45104: contig of 825 bp in length
 * 45105 45204: gap of 100 bp
 * 45205 46034: contig of 830 bp in length
 * 46035 46134: gap of 100 bp
 * 46135 46936: contig of 802 bp in length
 * 46937 47036: gap of 100 bp
 * 47037 47869: contig of 833 bp in length

```
Search completed: October 9, 2001, 12:12:46
Job time: 3881 sec
```

*	47870	47969:	gap of	100 bp	
*	47970	48770:	contig of 801 bp		In length
*	48771	48870:	gap of	100 bp	
*	48871	49668:	contig of 798 bp		In length
*	49669	49768:	gap of	100 bp	
*	49769	50560:	contig of 792 bp		In length
*	50561	50660:	gap of	100 bp	
*	51419	51418:	contig of 758 bp		In length
*	51519	51518:	gap of	100 bp	
*	52302	52301:	contig of 783 bp		In length
*	52402	52401:	gap of	100 bp	
*	52402	53189:	contig of 788 bp		In length
*	53190	53289:	gap of	100 bp	
*	53290	54081:	contig of 792 bp		In length
*	54082	54181:	gap of	100 bp	
*	54182	54984:	contig of 803 bp		In length
*	54885	55084:	gap of	100 bp	
*	55085	55863:	contig of 779 bp		In length
*	55864	55963:	gap of	100 bp	
*	55964	56778:	contig of 815 bp		In length
*	56779	56878:	gap of	100 bp	
*	56879	57657:	contig of 779 bp		In length
*	57658	57777:	gap of	100 bp	
*	57758	58554:	contig of 797 bp		In length
*	58555	58654:	gap of	100 bp	
*	58655	59472:	contig of 818 bp		In length
*	59473	59572:	gap of	100 bp	
*	59573	60360:	contig of 788 bp		In length
*	60361	60460:	gap of	100 bp	
*	60461	61240:	contig of 780 bp		In length
*	61241	61340:	gap of	100 bp	
*	61341	62135:	contig of 855 bp		In length
*	62146	62295:	gap of	100 bp	
*	62296	63086:	contig of 791 bp		In length
*	63087	63186:	gap of	100 bp	
*	63187	63978:	contig of 792 bp		In length
*	63979	64078:	gap of	100 bp	
*	64079	64874:	contig of 796 bp		In length
*	64875	64974:	gap of	100 bp	
*	64975	65754:	contig of 780 bp		In length

Query Match	71.28;	Score 17.8;	DB 64;	Length 68419;
Best Local Similarity	90.58;	Pred. No. 7.2e+02;		
Matches 19;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0

```
OY      5  gtgcatacgccagcatttcgat 25
          |||||  ||  |||||  |||
Db      28759 GTGCATCCCGCACCATTTTCGAT 28739
```

RESULT 45
AL359456_2
WPCOMMENT

Fragment Name	Begin	End	LOCUS	Accession
Sequence split into 6 fragments			AL359456	AL359456

AL359456_0	1	110000
AL359456_1	100001	210000
AL359456_2	200001	310000
AL359456_3	300001	410000
AL359456_4	400001	510000
AL359456_5	500001	593964

Continuation (3 of 6) of AL359456 from base 200001 (AL359456 Homo sapiens chromosome 10)

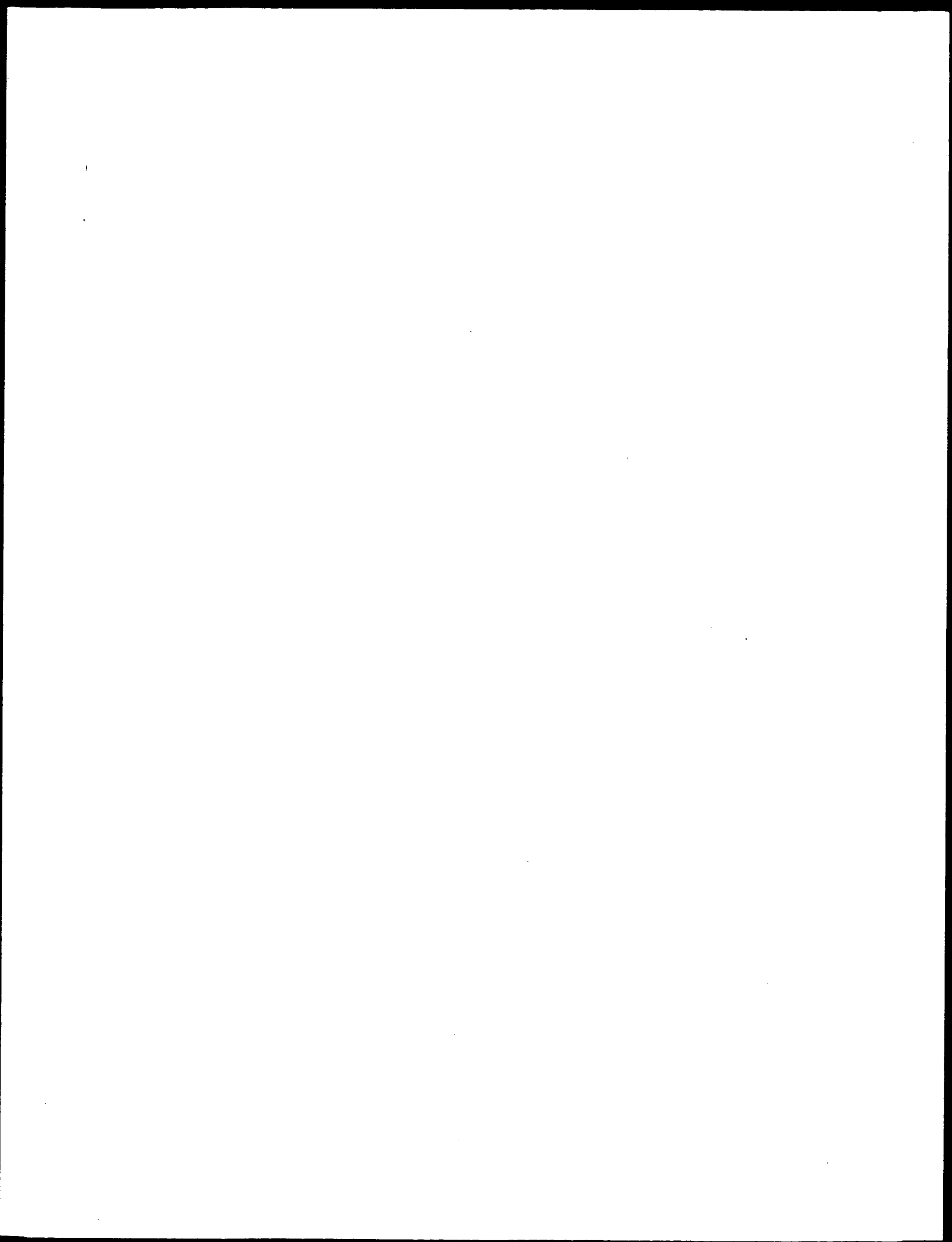
Query Match	71.2%;	Score 17.8;	DB 80;	Length 110000;
Best Local Similarity	90.5%;	Pred. No. 7e+02;		
Matches 19; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

```
QY      1  gcaggtgcacgccagcatlt 21
          |||||  |||||
Db 22667  GCAGGAACATCGCCAGCAATT 22687
```


Wed Oct 10 07:44:42 2001

us-09-396-196f-6.std.rge

Page 43



XX Example 2; Column 37-40; 34pp; English.

PS This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthetase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.

XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

QY Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 gcaggtgcatgcgcagcattcgat 25
 ||||||||||||||||||||||||||||
 84 gcaggtgcatgcgcagcattcgat 108

RESULT 2

XX AAN91329 standard; DNA; 1084 BP.

XX AAN91329;

XX 15-FEB-1990 (first entry)

XX E.coli Bio B gene.

XX E.coli; Bio B gene; biotin.

XX Escherichia coli.

XX Key Location/Qualifiers

XX CDS 24..1064

XX /*tag=a

XX GB2216530-A.

XX 11-OCT-1989.

XX 17-MAR-1989; 89GB-0006210.

XX 22-MAR-1988; 88GB-0006804.

XX 17-MAR-1989; 89GB-0006210.

XX (UKAG-) UK MIN. AGRIC. FISH.

XX Pearson BM, McKee RA;

XX WPI; 1989-295085/41. P-PSDB P91392

XX Table 3; page 33-4; 52pp; English.

XX The gene can be used in a plasmid for expression of enzymes of the biotin
 CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
 CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.

XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

QY Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 gcaggtgcatgcgcagcattcgat 25
 ||||||||||||||||||||||||||||
 107 gcaggtgcatgcgcagcattcgat 131

RESULT 3

XX AAN60496 standard; DNA; 1121 BP.

XX AAN60496;

XX 17-OCT-1991 (first entry)

XX Sequence encoding biotin synthesising enzyme.

XX Biotin synthetic enzyme; E.coli; desthiobiotin; ds.

XX Key Location/Qualifiers

XX CDS 42..1082

XX /*tag=a

XX JP61149091-A.

XX 07-JUL-1986.

XX 24-DEC-1984; 84JP-0272605.

XX 24-DEC-1984; 84JP-0272605.

XX (NIPS) NIPPON SODA KK.

XX WPI; 1986-216622/33.

XX P-PSDB; AAP60536.

XX Double stranded DNA encoding biotin synthesising enzyme -

XX comprises transformed mutant E.coli strain contg. cyclic doubled

XX stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

XX Disclosure; Page 534; 23pp; Japanese.

XX The sequence may be expressed by a transformed E.coli host, cultured

XX in a medium containing desthiobiotin.

XX Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

QY Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 0.013;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 gcaggtgcatgcgcagcattcgat 25
 ||||||||||||||||||||||||||||
 125 gcaggtgcatgcgcagcattcgat 149

XX AAN62386 standard; DNA; 5872 BP.

XX AAN62386;

XX 16-NOV-1994 (first entry)

XX Biotin-biosynthesis genes contg. plasmid PB030A-15/9.

XX Biotin; expression; enterobacteria; vitamin H; synthetase;

XX plasmid; PB030A-15/9; bioB; bioF; bioC; bioD; bioA;

XX promoter plac; biotin synthetase; KAPA synthetase;

XX 8-amino-7-oxononanoate synthetase; pimeoyl-CoA; DNB synthetase;

XX	dehbiobiotin synthetase: DAPA synthase;
KW	S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KM	seborrhea; dermatitis; ds.
XX	
OS	Escherichia coli DSM498.
XX	
FH	Location/Qualifiers
FT	Key
FT	promoter
FT	1..96
FT	/*tag= a
FT	/function= "promoter plac"
FT	/evidence= EXPERIMENTAL
FT	23..28
FT	/*tag= b
FT	/standard_name= "promoter plac"
FT	45..50
FT	/*tag= c
FT	/evidence= EXPERIMENTAL
FT	/standard_name= "promoter plac"
FT	105..109
FT	/*tag= d
FT	/evidence= EXPERIMENTAL
FT	/standard_name= "bioB RBS no. 9"
FT	117..1157
FT	/*tag= e
FT	/product= "biotin synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "bioB"
FT	/number= 1
FT	1141..1146
FT	/*tag= f
FT	/standard_name= "bioF RBS"
FT	1154..2311
FT	/*tag= g
FT	/EC_number= 2.3.1.47
FT	/product= "KAPA synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "bioF"
FT	/number= 2
FT	/standard_name= "8-amino-7-oxononanoate synthase"
FT	2284..2288
FT	/*tag= h
FT	/standard_name= "bioc RBS"
FT	2295..3050
FT	/*tag= i
FT	/function= "involved in pimeioyl-CoA synthesis"
FT	/product= "protein"
FT	/gene= "bioc"
FT	/number= 3
FT	3030..3033
FT	/*tag= j
FT	/standard_name= "biOD RBS"
FT	3043..3753
FT	/*tag= k
FT	/EC_number= 6.3.3.3
FT	/product= "DTB synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "biOD15"
FT	/number= 4
FT	/standard_name= "dethiobiotin synthase"
FT	3712..3750
FT	/*tag= l
FT	/note= "biOD15 substitution"
FT	3742..3746
FT	/*tag= m
FT	/standard_name= "bioA RBS"
FT	3750..5039
FT	/*tag= n
FT	/EC_number= 2.6.1.62
FT	/product= "DAPA synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "bioA"
FT	/number= 5
FT	/standard_name= "S-adenosyl-L-methionine: 8-amino

```

FT      7-oxononanoate aminotransferase"
FT      RBS
FT      5088..5093
FT      /*tag= O
FT      /standard_name= "ORFI RBS"
FT      CDS
FT      5098..5574
FT      /*tag= P
FT      /function= "unknown, involved in biotin synthesis"
FT      /product= "protein"
FT      /evidence= EXPERIMENTAL
FT      /gene= "ORFI"
FT      /number= 6
FT      5583..5644
FT      /*tag= q
FT      /standard_name= "rho-independent transcriptional
FT      terminator"
FT      stem_loop
FT      5583..5605
FT      /*tag= r
XX
XX      W09408023-A.
XX      14-APR-1994.
XX      PD
XX      01-OCT-1993; 93WO-EP02688.
XX      PF
XX      02-OCT-1992; 92CH-0003124.
XX      PR
XX      15-JUL-1993; 93CH-0002134.
XX      PA
XX      (LONZ ) LONZA AG.
XX      PI
XX      Birch O, Brass J, Fuhrmann M, Shaw N;
XX      WPI; 1994-135587/16.
XX      DR
XX      P-PSDB: AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
XX      PT
XX      Biotechnological biotin Prodn. using enterobacterial biotin-gene
XX      - providing vitamin H in high yield
XX      PS
XX      Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
XX      CC
XX      The sequence is derived from plasmid PB030A-15/9 contg. the
XX      CC      bioF, bioC, bioD and bio genes responsible for biosynthesis
XX      CC      of biotin, arranged in a transcription unit. Microorganisms
XX      CC      contg. these DNA fragments or plasmids may be used in the prodn.
XX      CC      of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
XX      CC      loss of appetite and tiredness.
XX      SO
XX      Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
XX
XX      Query Match 100.0%; Score 25; DB 15; Length 5872;
XX      Best Local Similarity 100.0%; Pred. NO. 0.017;
XX      Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
XX      QY
XX      1 gcaggtgcattgccagattcgat 25
XX      ||||||||||||||||||||
XX      Db 200 gcaggtgcattgccagattcgat 224
XX
XX      RESULT 5
XX      AAT68959
XX      ID AAT68959 standard; DNA; 3084 BP.
XX      AC
XX      AAT68959;
XX      AT
XX      05-AUG-1997 (first entry)
XX      DE
XX      Candida albicans chitin synthase (CHS1) gene.
XX      KW
XX      Chitin synthase; CHS1; yeast-associated disease; antifungal;
XX      KM
XX      fungicide; ds.
XX      OS
XX      Candida albicans.
XX

```

PN W09716540-A1.
 XX
 PD 09-MAY-1997.
 XX
 PF 01-NOV-1996; 96WO-US17459.
 XX
 PR 01-NOV-1995; 95US-0551437.
 XX
 PA (CHEM-) CHEMGENICS PHARM INC.
 XX
 PI Buiawa C, Gavrias V, Koltin Y, Riggle P, Winter K;
 DR WPI: 1997-272110/24.
 DR P-PSDB; AAM17057.
 XX
 FT New Candida albicans chitin synthase (CHS1) polypeptide and gene
 PT used to develop products for monitoring yeast-associated diseases
 PS and for developing antifungal drugs
 XX
 PS Claim 4; Page 32-36; 64pp; English.
 CC A 3084 bp genomic DNA sequence (AAT68959) from Candida albicans
 CC codes for a 1027-amino acid chitin synthase (CHS1) (AAM17057), an
 CC enzyme essential for cell wall synthesis and yeast cell growth, as
 CC well as for colonisation of tissues during infection in vivo.
 CC The full-length sequence was isolated using a plasmid rescue
 CC strategy and PCR. A novel maltose responsive promoter (see also
 CC AAT68960) was also isolated from C. albicans. The isolated CHS1 gene
 CC can be incorporated into an expression vector and used to produce
 CC CHS1 in host cells. The products can be used for monitoring
 CC yeast associated diseases and to develop antifungal drugs (e.g.
 CC anti-CHS1 antibodies and CHS1 antisense molecules).
 XX
 SO Sequence 3084 BP; 987 A; 550 C; 612 G; 935 T; 0 other;

Query Match 70.4%; Score 17.6; DB 18; Length 3084;
 Best Local Similarity 83.3%; Pred. No. 48;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 caggtgcatgcgcagcatttcgat 25
 |||||
 Db 1691 caggtgcatgcgcagcatttcgat 1714

RESULT 6
 ID AAT37079 standard; cDNA; 3617 BP.
 XX
 AC AAT37079;
 XX
 DT 16-OCT-1996 (first entry)
 XX
 DE AR301 fragment encoding MEI2 protein kinase PART1.
 XX
 KM Transformation; host plant; meiosis; spore formation; yeast; ds.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 452..3106
 FT /*tag= a
 FT /product= Protein kinase PART1
 XX
 PN JP08140686-A.
 XX
 PD 04-JUN-1996.
 XX
 PF 25-NOV-1994; 94JP-0315514.
 XX
 PR 25-NOV-1994; 94JP-0315514.
 XX
 PA (TOYT) TOYOTA JIDOSHA KK.

XX
 DR WPI: 1996-316320/32.
 DR P-PSDB; AAM00160.
 XX
 PT Gene coding for plant protein - useful in meiosis and spore
 PT formation
 XX
 PS Claim 1; Page 7-11; 26pp; Japanese.
 CC
 CC The sequences given in AAT37079-85 are cDNA sequences which were used
 CC to transform host plants. The proteins encoded by these sequences
 CC are involved in physical activities in plants such as meiosis, etc.
 CC in plant cells. The genes are useful for induction of meiosis and
 CC spore formation after introduction into a yeast incapable of meiosis
 CC and spore formation.
 XX
 SO Sequence 3617 BP; 1100 A; 700 C; 767 G; 1050 T; 0 other;

Query Match 70.4%; Score 17.6; DB 17; Length 3617;
 Best Local Similarity 83.3%; Pred. No. 49;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 caggtgcatgcgcagcatttcgat 25
 |||||
 Db 1757 caggtgcatgcgcagcatttcgat 1780

RESULT 7
 ID AAT39842 standard; DNA; 3186 BP.
 XX
 AC AAT39842;
 XX
 DT 29-SEP-1998 (first entry)
 XX
 DE Pseudomonas fluorescens Lemna gene.
 XX
 KM Pseudomonas; genetic engineering; biocontrol; plant; pathogenic;
 KM Rhizoctonia; Pythium; antifungal; pyrrolinidin; crop protection; ss.
 OS Pseudomonas fluorescens.
 XX
 FH Key Location/Qualifiers
 FT RBS 245..251
 FT /*tag= a
 FT /note= "potential ribosome binding site"
 FT 256..3006
 FT /*tag= b
 FT /product= "Lemna"
 XX
 PN W09824919-A1.
 XX
 PD 11-JUN-1998.
 XX
 PF 05-DEC-1997; 97WO-EP06815.
 XX
 PR 09-SEP-1997; 97US-0058304.
 PR 06-DEC-1996; 96US-0761258.
 XX
 PA (NOVS) NOVARTIS AG.
 XX
 PI Gaffney TD, Hill DS, Lam ST, Ligon JM, Stafford JM;
 PI Torckwitz NR;
 XX
 DR WPI: 1998-333337/29.
 XX
 PF Genetically modified Pseudomonas strains - useful to protect crop
 PF plants by controlling or inhibiting plant pathogen growth, e.g.
 PF growth of Rhizoctonia species
 XX
 PS Example 5; Page 63-66; 85pp; English.

CC A genetically engineered biocontrol strain of *Pseudomonas* has been
 CC developed that can control attacks on crop plants by pathogenic fungi,
 CC e.g. *Rhizoctonia* and *Pythium* and aggressively compete with indigenous
 CC bacteria and microflora in the plant rhizosphere. The strains can be
 CC included with agronomically acceptable carriers or chemical fungicides
 CC (e.g. metaxyl compounds) in biocontrol compositions. The strains or
 CC compositions can be applied to a plant/plant part to protect it from a
 CC plant pathogenic fungus, by controlling or inhibiting fungal growth.
 CC They can also be applied to the environment in which a plant pathogenic
 CC fungus will grow (e.g. soil) to similarly control or inhibit pathogen
 CC growth, or to seeds to protect plants developing from the seed from a
 CC plant pathogenic fungus. They are especially effective against
 CC *Rhizoctonia* and *Pythium* species which cause damping off in cotton.
 CC *Rhizoctonia* also infects many other crop species (e.g. beans and wheat),
 CC and no effective chemical fungicides are available. The present sequence
 CC represents the *Pseudomonas fluorescens* *lemA* gene used in an example from
 CC the present invention.

SQ Sequence 3186 BP; 646 A; 1091 C; 961 G; 488 T; 0 other;

Query Match

Best Local Similarity 68.8%; Score 17.2; DB 19; Length 3186;
 Pred. No. 74;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 caggtgcacgcacgacattcg 23

DB 2763 CAGGACGATGCGCCAGCATGTGCG 2742

RESULT 8

AAV31654/C standard; DNA; 3186 BP.

AC AAV31654;

DT 21-AUG-1998 (first entry)

DE Nucleotide sequence of the *lemA* gene.

XX Lem A gene; phytopathogenic fungi; *Rhizoctonia*; *Pythium* spp;

KW Pyrolnitrin; plant rhizospheres; ss.

XX *Pseudomonas fluorescens*.

Key Location/Qualifiers
 FH 256..3009
 FT /*tag= a
 FT /product= "Lem A protein"

US5756087-A.

26-MAY-1998.

06-DEC-1996; 96US-0761258.

06-DEC-1996; 96US-0761258.

(NOVS) NOVARTIS FINANCE CORP.

GAffney TD, Hill DS, Lam ST, Ligon JM, Torzewitz N;

WPI; 1998-321460/28.

Genetically modified *Pseudomonas fluorescens* strains - useful for
 biological control of phytopathogenic fungi

Example 5; Columns 43-48; 35pp; English.

CC This is the nucleotide sequence of the *Pseudomonas fluorescens*
 CC *lemA* gene, which is used in the method of the invention to produce
 CC modified *P. fluorescens* strains. The modified strains are useful for
 CC biological control of phytopathogenic fungi, especially *Rhizoctonia*

CC or *Pythium* spp. The strains produce high levels of pyrolnitrin and
 CC are able to compete with indigenous bacteria and microflora in plant
 CC rhizospheres.

SQ Sequence 3186 BP; 646 A; 1091 C; 961 G; 488 T; 0 other;

Query Match

Best Local Similarity 68.8%; Score 17.2; DB 19; Length 3186;
 Pred. No. 74;

Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 caggtgcacgcacgacattcg 23

DB 2763 CAGGACGATGCGCCAGCATGTGCG 2742

RESULT 9

AAV31654/C standard; DNA; 3186 BP.

AC AAV31654;

DT 15-NOV-1999 (first entry)

DE *P. fluorescens* *lemA* gene sequence.

XX Biocontrol strain; *Pseudomonas fluorescens*; pyrolnitrin; crop plant;
 KW fungal attack; pathogenic; *Rhizoctonia*; *Pythium*; *lemA* gene; ss.

XX *Pseudomonas fluorescens*.

Key Location/Qualifiers
 FH 245..251
 FT /*tag= a
 FT /note= "potential ribosome binding site"

FT CDS 256..3006

FT /*tag= b
 FT /product= "LemA"

US5955348-A.

21-SEP-1999.

25-NOV-1997; 97US-0977306.

25-NOV-1997; 97US-0977306.

(NOVS) NOVARTIS AG.

GAffney TD, Hill DS, Ligon JM, Stafford JM, Torzewitz NR;

WPI; 1999-539577/45.

Strains of *Pseudomonas fluorescens* with enhanced pyrolnitrin
 production useful for the biocontrol of plant pathogens

Example 5; Columns 51-54; 33pp; English.

CC The invention provides novel biocontrol strains of *P. fluorescens* or
 CC their pyrolnitrin producing progeny. These strains of *Pseudomonas* have
 CC been genetically engineered to have enhanced biocontrol properties.
 CC (see AAV31654 for the identification names and NRRL accession numbers
 CC for the different biocontrol strains of *P. fluorescens*). The strains of
 CC *P. fluorescens* may be used to control pathogenic attack on crop plants,
 CC especially by fungi such as *Rhizoctonia* or *Pythium*. The strains produce
 CC enhanced amounts of pyrolnitrin and aggressively compete with
 CC indigenous bacteria and microflora existing in the rhizosphere of the
 CC plant. The present sequence represents the *lemA* gene sequence.

SQ Sequence 3186 BP; 646 A; 1091 C; 961 G; 488 T; 0 other;

Query Match

Best Local Similarity 68.8%; Score 17.2; DB 20; Length 3186;

Best Local Similarity 86.4%; Pred. No. 74;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 caggtgcacgcagcattccg 23
||||| ||||| ||||| |||
Db 2763 CAGGAGCATGCCAGCATGTCG 2742

RESULT 10

AAZ53592
ID AAZ53592 standard; DNA: 669 BP.

XX AAZ53592;

XX 21-MAR-2000 (first entry)

DE Neisseria gonorrhoeae ORF 277 partial DNA sequence SEQ ID NO:1133.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KM antibacterial; gene therapy; ds.

XX OS Neisseria gonorrhoeae.

XX PN W09957280-A2.

XX PD 11-NOV-1999.

XX PE 30-APR-1999; 99WO-US09346.

XX PF 01-MAY-1998; 98US-0083758.

XX PR 31-JUL-1998; 98US-0094869.

XX PR 02-SEP-1998; 98US-0098994.

XX PR 02-SEP-1998; 98US-0099062.

XX PR 09-OCT-1998; 98US-0103749.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 25-FEB-1999; 99US-0121528.

XX PA (CHIR) CHIRON CORP.

XX PA (GENO-) INST GENOMIC RES.

XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

XX PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

XX PI Tettelin H, Venter JC;

XX DR WPI: 2000-062150/05.

XX DR P-PSDB: AAY74830.

XX DR Novel Neisserial polypeptides predicted to be useful antigens for

XX PT vaccines and diagnostics -

XX PS Claim 7; Page 630; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941

XX CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides

XX CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent

XX CC PCR primers used in the exemplification of the present invention. The

XX CC polypeptides, the polynucleotides, antibodies and compositions of

XX CC the invention can be used as vaccines, as diagnostic reagents, and as

XX CC immunogenic compositions. The polypeptides can be used in the

XX CC manufacture of medicaments for treating or preventing infection due to

XX CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the

XX CC presence of Neisseria bacteria, or to raise antibodies. They may also

XX CC be used to screen for agonists or antagonists, which may themselves

XX CC have use as antibacterial agents. The polynucleotides of the invention

XX CC may also be used in gene therapy protocols.

XX SQ Sequence 669 BP; 133 A; 149 C; 219 G; 168 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 669;
Best Local Similarity 80.0%; Pred. No. 71;

Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcagcattccg 25
||||| ||||| ||||| |||||
Db 81 gcaggtgcacgcagcattccg 105

RESULT 11

AAZ53594
ID AAZ53594 standard; DNA: 759 BP.

XX AAZ53594;

XX 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 277 partial DNA sequence SEQ ID NO:1137.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KM antibacterial; gene therapy; ds.

XX OS Neisseria meningitidis.

XX PN W09957280-A2.

XX PD 11-NOV-1999.

XX PE 30-APR-1999; 99WO-US09346.

XX PF 01-MAY-1998; 98US-0083758.

XX PR 31-JUL-1998; 98US-0094869.

XX PR 02-SEP-1998; 98US-0098994.

XX PR 02-SEP-1998; 98US-0099062.

XX PR 09-OCT-1998; 98US-0103749.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 25-FEB-1999; 99US-0121528.

XX PA (CHIR) CHIRON CORP.

XX PA (GENO-) INST GENOMIC RES.

XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

XX PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

XX PI Tettelin H, Venter JC;

XX DR WPI: 2000-062150/05.

XX DR P-PSDB: AAY74832.

XX DR Novel Neisserial polypeptides predicted to be useful antigens for

XX PT vaccines and diagnostics -

XX PS Claim 7; Page 631; 1453pp; English.

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941

XX CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides

XX CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent

XX CC PCR primers used in the exemplification of the present invention. The

XX CC polypeptides, the polynucleotides, antibodies and compositions of

XX CC the invention can be used as vaccines, as diagnostic reagents, and as

XX CC immunogenic compositions. The polypeptides can be used in the

XX CC manufacture of medicaments for treating or preventing infection due to

XX CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the

XX CC presence of Neisseria bacteria, or to raise antibodies. They may also

XX CC be used to screen for agonists or antagonists, which may themselves

XX CC have use as antibacterial agents. The polynucleotides of the invention

XX CC may also be used in gene therapy protocols.

XX SQ Sequence 759 BP; 138 A; 173 C; 261 G; 187 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 759;
Best Local Similarity 80.0%; Pred. No. 73;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcgcagcattcgat 25
 ||||| ||||| |||||
 Db 243 gcaggtacatgcgcagcgttcgat 267

RESULT 12
 AAX99364
 ID AAX99364 standard; DNA; 6386 BP.
 AC AAX99364;
 XX
 DT 09-NOV-1999 (first entry)
 XX
 DE Maize shrunken-1 (Sh1) gene sequence.
 XX
 KM Gene expression; shrunken-1 gene; Sh1 gene; maize; foreign gene; ds.
 XX
 OS Zea mays.

Location/Qualifiers
 FT key 131..182
 FT exon /*tag= a
 FT /number= 1
 FT 183..1210
 FT /*tag= b
 FT /number= 1
 FT 1211..1324
 FT /*tag= c
 FT /number= 2
 FT 1325..1827
 FT /*tag= d
 FT /number= 2
 FT 1828..1948
 FT /*tag= e
 FT /number= 3
 FT 1949..2040
 FT /*tag= f
 FT /number= 3
 FT 2041..2187
 FT /*tag= g
 FT /number= 4
 FT 2188..2268
 FT /*tag= h
 FT /number= 4
 FT 2269..2460
 FT /*tag= i
 FT /number= 5
 FT 2461..2604
 FT /*tag= j
 FT /number= 5
 FT 2605..2728
 FT /*tag= k
 FT /number= 6
 FT 2729..2821
 FT /*tag= l
 FT /number= 6
 FT 2822..3038
 FT /*tag= m
 FT /number= 7
 FT 3039..3255
 FT /*tag= n
 FT /number= 7
 FT 3256..3351
 FT /*tag= o
 FT /number= 8
 FT 3352..3446
 FT /*tag= p
 FT /number= 8
 FT 3447..3620
 FT /*tag= q
 FT /number= 9
 FT 3621..3701
 FT intron

FT /*tag= r
 FT /number= 9
 FT 3702..3818
 FT /*tag= s
 FT /number= 10
 FT 3819..3911
 FT /*tag= t
 FT /number= 10
 FT 3912..4078
 FT /*tag= u
 FT /number= 11
 FT 4079..4157
 FT /*tag= v
 FT /number= 11
 FT 4158..4381
 FT /*tag= w
 FT /number= 12
 FT 4382..4516
 FT /*tag= x
 FT /number= 12
 FT 4517..4835
 FT /*tag= y
 FT /number= 13
 FT 4836..4767
 FT /*tag= z
 FT /number= 13
 FT 4768..5212
 FT /*tag= aa
 FT /number= 14
 FT 5213..5371
 FT /*tag= ab
 FT /number= 14
 FT 5372..5510
 FT /*tag= ac
 FT /number= 15
 FT 5511..5635
 FT /*tag= ad
 FT /number= 15
 FT 5636..5917
 FT /*tag= ae
 FT /number= 16
 FT exon

US5955330-A.
 PN 21-SEP-1999.
 XX
 PD
 XX
 PF 07-JUN-1995; 95US-0483376.
 XX
 PR 18-MAY-1989; 89US-0353854.
 PR 05-FEB-1992; 92US-0830956.
 PR 04-AUG-1993; 93US-0102115.
 PR 07-APR-1995; 95US-0418540.
 PR 07-JUN-1995; 95US-0483376.

(RESE) RESEARCH CORP TECHNOLOGIES INC.
 PA
 XX
 PI Clancy MA, Ferl RJ, Hannah LC, Vasil IK, Vasil V;
 XX
 DR WPI: 1999-539573/45.
 XX
 PT
 XX
 PS

Increasing the expression of a gene in a plant using an intron

Claim 6: Fig 1A-G: 32pp: English.

The invention relates to a method for increasing the expression rate of a gene in a plant cell using the first intron of the shrunken-1 (Sh1) locus of maize. The method comprises (a) inserting an intron into the gene comprising a 5' and 3' splice site of the Sh1 first intron, where the intron is inserted between a transcription start site and a translation start site of the gene to form an intron modified gene; and (b) introducing the modified gene into a plant cell so that the gene is expressed. The method is useful for increasing the expression of foreign genes in plant cells. The present sequence represents the maize Sh1

CC presence of Neisserial bacteria or of antibodies raised to Neisserial
 CC bacteria. Computers, computer memory, computer storage medium or computer
 CC databases can be used in a search to identify open reading frames (ORFs)
 CC or coding sequences within the NMB genome. The DNA sequences provide
 CC further opportunities to find antigenic or immunogenic proteins which are
 CC more effective in vaccines than the outer membrane proteins currently
 CC used.

SQ Sequence 349980 BP; 86771 A; 92803 C; 86340 G; 84066 T; 0 other;

Query Match 68.0%; Score 17; DB 21; Length 349980;
 Best Local Similarity 80.0%; Pred. No. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gcaggtgcacgcacgattccat 25
 |||||
 Db 222215 GCAGGTGCACGCCCATCATTTCCAT 222191

RESULT 15
 ID AAA81490/C
 AC AAA81490 standard; DNA; 1437668 BP.

XX AAA81490;
 XX
 DT 04-DEC-2000 (first entry)
 XX

DE N. meningitidis B full length genome DNA sequence SEQ ID NO:1068.

XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW Meningococcus B; MenB; ds.

XX Neisseria meningitidis.

OS WO200022430-A2.

PN 20-APR-2000.

XX 08-OCT-1999; 99WO-US23573.

XX 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.

XX (CHIR) CHIRON CORP.

PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;

PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappelli R, Pizsa M;

PI WPI; 2000-318079/27.

XX Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -

XX Claim 7; Page 866-1272; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences. AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins. AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could

CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

SQ Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

Query Match 68.0%; Score 17; DB 21; Length 1437668;
 Best Local Similarity 80.0%; Pred. No. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 gcaggtgcacgcacgattccat 25
 |||||
 Db 1392603 GCAGGTGCACGCCCATCATTTCCAT 1392579

RESULT 16
 ID AAV69543/C
 AC AAV69543 standard; DNA; 643 BP.

XX AAV69543;
 XX

DT 15-MAR-1999 (first entry)
 XX

DE P. neopolydactyla type I polyketide synthase DNA #2.

XX Type I polyketide synthase; soil; lichen; antibiotic biosynthesis;
 KM humus; therapeutic; immunosuppressor; antitumor agent; pathogen;
 KW genetic diversity; ss.

OS Peltigera neopolydactyla.

PN WO9853097-A2.

XX 26-NOV-1998.

XX 21-MAY-1998; 98WO-CA00488.

XX 22-MAY-1997; 97US-0861774.

XX (TERR-) TERRAGEN DIVERSITY INC.

XX Miao VPW, Seow KT, Waters B, Yap WH;

XX WPI; 1999-070158/06.

DR P-PSDB; AAW82695.

XX New degenerate primers - used for recovering antibiotic biosynthetic
 PT DNA from soil/lichen material

XX Claim 17; Page 57; 98pp; English.

XX This sequence encodes a Type I polyketide synthase from the lichen
 CC Peltigera neopolydactyla. This protein is used in a method for the
 CC recovery of antibiotic biosynthetic DNA from humic materials or lichen.
 CC The PCR products of the invention have the potential to be used as
 CC therapeutic molecules including antibiotics, immunosuppressors and
 CC antitumor agents. The method allows access to the reservoir of genetic
 CC diversity in soil pathogenic micro-organisms, in order to find new
 CC antibiotics. It also allows access to novel biosynthetic genes/enzymes
 CC that can be used to produce antibiotics or produce specific compounds,
 CC enzymatically, in vitro.

SQ Sequence 643 BP; 175 A; 143 C; 156 G; 169 T; 0 other;

```

Query Match      67.2%; Score 16.8; DB 20; Length 643;
Best Local Similarity 90.0%; Pred. No. 88;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      2 caggtcgcgcgcagcattt 21
        ||||| |||||
Db      53 CAGTGCCTTGCACGACATTT 34

RESULT 17
AAT64616
ID AAT64649 standard; DNA; 65 BP.
XX
AC AAT64649;
XX
DF 17-JAN-1998 (first entry)
XX
DE G-CSF receptor agonist gene primer 125for.
XX
KW Granulocyte colony stimulating factor receptor; agonist; G-CSF;
KW haematopoietic disorder; neutropenia; bone marrow suppression;
KW stem cell expansion; gene therapy; circular permutation;
KW polymerase chain reaction; PCR; primer; ss.
XX
OS Synthetic.
XX
PN WO9712977-A1.
XX
PI 10-APR-1997.
XX
PE 04-OCT-1996; 96WO-US15935.
XX
PR 05-OCT-1995; 95US-0004832.
XX
PA (SEAR ) SEARLE & CO G D.
XX
PI Braford-Goldberg SR, Feng Y, Klein BK, McKearn JP;
PI McWherter CA, Zurfluh LL;
XX
DR WPI; 1997-244718/22.
XX
PT Modified human granulocyte colony stimulating factor - useful as
PT G-CSF receptor agonist for treating haematopoietic disorders, e.g.
PT neutropenia or bone marrow suppression
XX
PS Example 25; Page 31; 186pp; English.
XX
CC This synthetic oligonucleotide comprises primer 125for that was
CC used in the construction of novel genes (see AAT64612 and AAT64616)
CC encoding circularly permuted variants (see AAM15040 and AAM15044) of
CC human granulocyte colony stimulating factor (G-CSF) that act as
CC G-CSF receptor agonists. Such receptor agonists can be used in
CC methods for stimulating production of haematopoietic cells, ex
CC vivo expansion of stem cells, treatment of haematopoietic disorders
CC and human gene therapy.
XX
SQ Sequence 65 BP; 16 A; 19 C; 13 G; 17 T; 0 other;

Query Match      66.4%; Score 16.6; DB 18; Length 65;
Best Local Similarity 82.6%; Pred. No. 76;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      3 aggtgcgcgcgcagcattgat 25
        ||||| |||||
Db      40 aggtgcgcgcgcagcattgat 62

RESULT 18
AAT64616
ID AAT64616 standard; DNA; 522 BP.
XX

AC AAT64616;
XX
DF 18-JAN-1998 (first entry)
XX
DE G-CSF receptor agonist gene in pMON25186.
XX
KW Granulocyte colony stimulating factor; receptor agonist; G-CSF;
KW haematopoietic disorder; neutropenia; bone marrow suppression;
KW stem cell expansion; gene therapy; circular permutation; pMON25186;
XX
OS Chimeric Homo sapiens.
XX
OS Chimeric synthetic.
XX
FH Key Location/Qualifiers
FT mat_peptide 7..519
FT /*tag= a
XX
PN WO9712977-A1.
XX
PI 10-APR-1997.
XX
PE 04-OCT-1996; 96WO-US15935.
XX
PR 05-OCT-1995; 95US-0004832.
XX
PA (SEAR ) SEARLE & CO G D.
XX
PI Braford-Goldberg SR, Feng Y, Klein BK, McKearn JP;
PI McWherter CA, Zurfluh LL;
XX
DR WPI; 1997-244718/22.
DR P-PSDB; AAM15044.
XX
PT Modified human granulocyte colony stimulating factor - useful as
PT G-CSF receptor agonist for treating haematopoietic disorders, e.g.
PT neutropenia or bone marrow suppression
XX
PS Claim 7; Page 171; 186pp; English.
XX
CC This DNA sequence comprises a new N-terminus/C-terminus gene in
CC pMON25186 that codes for a specific example of a granulocyte colony
CC stimulating factor (G-CSF) receptor agonist (see AAM15044) comprising
CC circularly permuted human G-CSF in which a C-terminal fragment of
CC G-CSF is linked to a truncated N-terminal fragment via a peptide
CC linker, the breakpoint being at residues 125/126. It was
CC constructed using an overlapping PCR primer approach (see AAT64649-
CC 50), resulting in 19 codon changes at the 5' end of the gene. The
CC receptor agonist can be expressed in transformed or transfected
CC host cells (e.g. E. coli). Claimed receptor agonists (see AAM15034-
CC 38) can be used in claimed methods for stimulating production of
CC haematopoietic cells in a patient, for selective ex vivo expansion
CC of stem cells, for treatment of a patient having a haematopoietic
CC disorder, and in methods of human gene therapy.
XX
SQ Sequence 522 BP; 94 A; 168 C; 148 G; 112 T; 0 other;

Query Match      66.4%; Score 16.6; DB 18; Length 522;
Best Local Similarity 82.6%; Pred. No. 11e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy      3 aggtgcgcgcgcagcattgat 25
        ||||| |||||
Db      33 aggtgcgcgcgcagcattgat 55

RESULT 19
AAT64612
ID AAT64612 standard; DNA; 531 BP.
XX
AC AAT64612;
XX

```

DT 18-JAN-1998 (first entry)

XX G-CSF receptor agonist gene in pMON25182.

XX Granulocyte colony stimulating factor; receptor agonist; G-CSF;

XX haematopoietic disorder; neutropenia; bone marrow suppression;

XX stem cell expansion; gene therapy; circular permutation; pMON25182;

XX ss.

XX Chimeric Homo sapiens.

OS Chimeric synthetic.

XX

XX Key Location/Qualifiers

FT mat_peptide 7..528

FT /*tag= a

XX

XX MO9712977-A1.

XX

XX 10-APR-1997.

XX

XX 04-OCT-1996; 96WO-US15935.

XX

XX 05-OCT-1995; 95US-0004832.

XX

XX (SEAR) SEARLE & CO G D.

XX

XX Braford-Goldberg SR, Feng Y, Klein BK, McKearn JP;

PI McWhorter CA, Zurfluh LL;

XX

XX WPI: 1997-244718/22.

XX P-PSDB; AAM15040.

XX

XX Modified human granulocyte colony stimulating factor - useful as

PT G-CSF receptor agonist for treating haematopoietic disorders, e.g.

PT neutropenia or bone marrow suppression

XX

XX Claim 7; Page 170; 186pp; English.

XX

XX This DNA sequence comprises a new N-terminus/C-terminus gene in

CC pMON25182 that codes for a specific example of a granulocyte colony

CC stimulating factor (G-CSF) receptor agonist (see AAM15040) comprising

CC circularly permuted human G-CSF in which the N-terminus corresponds

CC to amino acid residue 126 of a generic G-CSF receptor agonist (see

CC AAM15025) and the C-terminus to residue 125. It was constructed

CC using an overlapping PCR primer method (see also AAT64649-50),

CC resulting in 19 codon changes at the 5' end of the gene. The

CC receptor agonist can be expressed in transformed or transfected

CC host cells (e.g. E. coli). Claimed receptor agonists (see AAM15034-

CC 38) can be used in claimed methods for stimulating production of

CC haematopoietic cells in a patient, for selective ex vivo expansion

CC of stem cells, for treatment of a patient having a haematopoietic

CC disorder, and in methods of human gene therapy.

XX

XX Sequence 531 BP; 98 A; 176 C; 146 G; 111 T; 0 other;

XX

XX Query Match 66.4%; Score 16.6; DB 18; Length 531;

XX Best Local Similarity 82.6%; Pred. No. 1.1e+02;

XX Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

XX

XX 3 aggtgcatgcgcagattcgat 25

XX ||||| ||||| ||

XX Db 33 aggtgcatgcgcagattcgat 55

XX

XX RESULT 20

XX AAX04354

XX ID AAX04354 standard; DNA: 2074 BP.

XX

XX AAX04354;

XX

XX 13-APR-1999 (first entry)

XX

DE Human secreted protein gene 44 clone HETF05.

XX

XX Human; secreted protein; fusion protein; gene therapy; protein therapy;

KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukemia;

KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;

KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;

KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;

KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;

KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;

KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX

XX Homo sapiens.

XX

XX WO9856804-A1.

XX

XX 17-DEC-1998.

XX

XX 11-JUN-1998; 98WO-US12125.

XX

XX 02-OCT-1997; 97US-0061060.

XX 13-JUN-1997; 97US-0049547.

XX 13-JUN-1997; 97US-0049548.

XX 13-JUN-1997; 97US-0049549.

XX 13-JUN-1997; 97US-0049550.

XX 13-JUN-1997; 97US-0049606.

XX 13-JUN-1997; 97US-0049607.

XX 13-JUN-1997; 97US-0049608.

XX 13-JUN-1997; 97US-0049609.

XX 13-JUN-1997; 97US-0049610.

XX 13-JUN-1997; 97US-0049611.

XX 13-JUN-1997; 97US-0050561.

XX 13-JUN-1997; 97US-0052989.

XX 13-JUN-1997; 97US-0051919.

XX 08-JUL-1997; 97US-0051919.

XX 18-AUG-1997; 97US-0055984.

XX 12-SEP-1997; 97US-0058668.

XX 12-SEP-1997; 97US-0058669.

XX 12-SEP-1997; 97US-0058670.

XX 12-SEP-1997; 97US-0058750.

XX 12-SEP-1997; 97US-0058971.

XX 12-SEP-1997; 97US-0058972.

XX 12-SEP-1997; 97US-0058975.

XX 02-OCT-1997; 97US-0060834.

XX 02-OCT-1997; 97US-0060841.

XX 02-OCT-1997; 97US-0060844.

XX 02-OCT-1997; 97US-0060865.

XX 02-OCT-1997; 97US-0061059.

XX

XX (HUMA-) HUMAN GENOME SCI INC.

XX

XX Brewer LA, Edner R, Ferlie AM, Feng P, Greene JM, Lafleur DW;

PI Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM, Shi Y, Young P;

PI Yu GL;

XX

XX WPI: 1999-080881/07.

XX P-PSDB; AAW78169.

XX

XX New isolated human genes and the secreted polypeptides they encode -

PT useful for diagnosis and treatment of e.g. cancers, neurological

PT disorders, immune diseases, inflammation or blood disorders

XX

XX Claim 1; Page 210-212; 380pp; English.

XX

XX This sequence represents a nucleic acid molecule which encodes a

CC secreted human protein. The gene number, and the clone it is derived

CC from, are detailed in the descriptor line. The gene can be used to

CC generate fusion proteins by linking to the gene to a human immunoglobulin

CC Fc portion (e.g. AAX04302) for increasing the stability of the fused

CC protein as compared to the human protein only.

CC The invention relates to 86 novel genes and their fragments (nucleic

CC acid sequences: AAX04311-X04410; amino acid sequences AAW78126-W78225)

CC which are useful for preventing, treating or ameliorating medical

CC conditions e.g. by protein or gene therapy. Also, pathological


```

XX AC AAX13070;
XX XX
XX DT 19-MAR-1999 (first entry)
XX XX
XX DE Enterococcus faecalis genome contig SEQ ID NO:133.
XX XX
XX KM Enterococcus faecalis; contig; detection; Enterococcal infection;
XX KM vaccine; attenuation; computer readable medium; ds.
XX OS Enterococcus faecalis.
XX PN MO9850555-A2.
XX PD 12-NOV-1998.
XX PF 04-MAY-1998; 98WO-US08985.
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
XX PR 16-MAY-1997; 97US-0046655.
XX PA (HUMA-) HUMAN GENOME SCL INC.
XX PI Barash SC, Dillon PU, Kunsch CA;
XX DR WPI; 1999-045171/04.
XX PT New isolated Enterococcus faecalis polynucleotides and polypeptides
XX PT - used to develop products for the detection of Enterococcus and for
XX PT use in vaccines for prevention or attenuation of Enterococcus
XX PT infection.
XX PS Claim 1; Page 799-804; 2084pp; English.
XX CC A computer readable medium has been developed which has recorded on it
XX CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX CC AAX12938 to AAX13919 represent these nucleotide sequences which are
XX CC primary nucleotide sequences, also known as contigs. The computer-based
XX CC system can identify fragments of the Enterococcus faecalis genome with
XX CC commercial importance. The products can be used to detect the presence
XX CC of Enterococcus faecalis in samples. They can also be used for
XX CC diagnosing Enterococcal infection in an animal and monitoring
XX CC progression of disease, and for identifying agents which can be used to
XX CC modulate the growth or pathogenicity of Enterococcus faecalis, or
XX CC another related organism, in vivo or in vitro. In particular the
XX CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
XX CC can be used in vaccines to prevent or attenuate an Enterococcal
XX CC infection.
XX SQ Sequence 10391 BP; 3471 A; 1709 C; 2240 G; 2962 T; 9 other:

Query Match 66.4%; Score 16.6; DB 20; Length 10391;
Best Local Similarity 82.6%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 2 cagagtcacgcgcagcattcga 24
DB 4469 CAGGTGATCGCCATCATTTCTA 4447

RESULT 23
AAQ46806/c
ID AAQ46806 standard; DNA: 29879 BP.
XX AC AAQ46806;
XX DT 22-DEC-1993 (first entry)
XX DE eryA region of S. erythraea chromosome.
XX KM Saccharopolyspora erythraea; eryA; biosynthesis; polyketide; module;

```

```

XX KM erythromycin; condensation; elongation; acyl chain growth;
XX KM gene replacement; SS.
XX OS Saccharopolyspora erythraea.
XX FH Key Location/Qualifiers
XX FT CDS 3..10199
XX FT /*tag= a
XX FT /note= "ORF 1"
XX FT 10218..20921
XX FT /*tag= b
XX FT /note= "ORF 2"
XX FT 20922..29879
XX FT CDS /*tag= c
XX FT /note= "ORF 3"
XX PN WO9313663-A.
XX PD 22-JUL-1993.
XX PR 17-JAN-1992; 92WO-US00427.
XX PR 17-JAN-1992; 92WO-US00427.
XX PA (ABBO ) ABBOTT LAB.
XX PI Donadio S, Katz L, McAlpine JB;
XX DR WPI; 1993-242804/30.
XX DR P-PSDB; AAR44430-32.
XX PT Biosynthesis of specific polyketide analogues esp. erythromycin
XX PT cpds. - by introducing altered biosynthetic gene-contg. DNA into
XX PT microorganisms
XX PS Claim 27; Fig 2; 133pp; English.
XX CC This sequence represents a fragment of the Saccharopolyspora erythraea
XX CC genome, designated eryA. The polypeptides encoded by this region
XX CC are involved in the biosynthesis of the polyketide segment of
XX CC erythromycin. eryA is organised in modules and each module takes care
XX CC of one condensation step. The precise succession of elongation steps
XX CC is dictated by the genetic order of the modules. This fragment may be
XX CC specifically altered such that novel polyketide molecules of desired
XX CC structure are produced. Three types of alteration may be produced:
XX CC those inactivating a single function in a module which does not arrest
XX CC acyl chain growth; those inactivating a single function in a module
XX CC which does affect chain growth; and those affecting an entire module.
XX CC The mutations may be introduced by gene replacement.
XX SQ Sequence 29879 BP; 3672 A; 10596 C; 11545 G; 4066 T; 0 other:

Query Match 66.4%; Score 16.6; DB 14; Length 29879;
Best Local Similarity 82.6%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 gcaggtcactgcgcagcattcg 23
DB 14949 GCGGGTCCATCCGCCAGCGCTTCG 14927

RESULT 24
AAF26295/c
ID AAF26295 standard; DNA: 2048 BP.
XX AC AAF26295;
XX DT 02-MAY-2001 (first entry)
XX DE Pseudomonas sp Type II (xcp) secretion system DNA ORF04962a.
XX KM Type II secretion system; xcp secretion system; ORF04962a; ds.

```


PR 25-FEB-1999; 99US-0121528.
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 DR WPI: 2000-062150/05.
 DR P-PSDB: AAY75579.
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 7; Page 1247; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 645 BP; 168 A; 147 C; 170 G; 160 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 645;
 Best Local Similarity 85.7%; Pred. No. 1.7e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 5 gtgcatgcgcagcattcgat 25
 ||||| ||||| ||||| |||||
 Db 160 gtacattgcacatcattcgat 180

RESULT 27
 AA254341
 ID AA254341 standard; DNA; 645 BP.
 XX
 AC AA254341;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 767 partial DNA sequence SEQ ID NO:2631.
 XX
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KM antibacterial; gene therapy; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.

XX
 XX (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 DR WPI: 2000-062150/05.
 DR P-PSDB: AAY75579.
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 PS Claim 7; Page 1247-1248; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SQ Sequence 645 BP; 170 A; 146 C; 170 G; 159 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 645;
 Best Local Similarity 85.7%; Pred. No. 1.7e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 5 gtgcatgcgcagcattcgat 25
 ||||| ||||| ||||| |||||
 Db 160 gtacattgcacatcattcgat 180

RESULT 28
 AA253354/C
 ID AA253354 standard; DNA; 900 BP.
 XX
 AC AA253354;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 158 partial DNA sequence SEQ ID NO:657.
 XX
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KM antibacterial; gene therapy; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.

PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR P-PSDB: AAY74592.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX
 PS Claim 7: Page 452; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC have use as antibacterial agents, or to raise antibodies. They may also
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SO Sequence 900 BP; 226 A; 231 C; 232 G; 211 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 900;
 Best Local Similarity 85.7%; Pred. No. 1.8e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 gtgcacgcacgacattcgat 25
 ||||| ||||| ||||| ||||| |||||
 Db 264 GTGCACCTGCCAGCATTTCCGT 244

RESULT 29
 AA253353/C
 ID AA253353 standard; DNA: 900 BP.
 XX
 AC AA253353;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 158 partial DNA sequence SEQ ID NO:659.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.

PA (GENO-) INST GENOMIC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR P-PSDB: AAY74593.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX
 PS Claim 7: Page 453; 1453pp; English.
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC have use as antibacterial agents, or to raise antibodies. They may also
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX
 SO Sequence 900 BP; 225 A; 234 C; 232 G; 209 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 900;
 Best Local Similarity 85.7%; Pred. No. 1.8e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 gtgcacgcacgacattcgat 25
 ||||| ||||| ||||| ||||| |||||
 Db 264 GTGCACCTGCCAGCATTTCCGT 244

RESULT 30
 AA253353/C
 ID AA253353 standard; DNA: 912 BP.
 XX
 AC AA253353;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria gonorrhoeae ORF 158 partial DNA sequence SEQ ID NO:655.
 XX
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
 KW antibacterial; gene therapy; ds.
 XX
 OS Neisseria gonorrhoeae.
 XX
 PN WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 XX 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcaggtgcacgcgcagcattt 21
 ||| ||| || ||||| |||||
 DB 1235 GCATGTGAATTGCCAGCATTT 1215

RESULT 35
 AAC77013/C
 ID AAC77013 standard; DNA: 3666 BP.
 XX
 AC AAC77013;
 XX
 DT 10-AUG-1999 (first entry)
 XX
 DE Mouse Sel-1L splice variant coding sequence.
 XX
 KW Sel-1L; Sel-1 like protein; Hip-1; Alzheimer's disease; diabetes; cancer;
 KW insulin dependent diabetes mellitus; pancreatic cancer; stroke; therapy;
 KW vascular dementia; Parkinson's disease; coronary heart disease;
 KW fat metabolism; cholesterol metabolism; ss.
 XX
 OS Mus sp.
 XX
 PN MO9927088-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 19-NOV-1998; 98WO-CA01058.
 XX
 PR 28-JUL-1998; 98US-0123549.
 PR 19-NOV-1997; 97US-0066140.
 XX
 PA (MOUN) MOUNT SINAI HOSPITAL.
 XX
 PI Bernstein A, Donoviel D;
 XX
 DR WPI: 1999-357833/30.
 DR P-PSDB; AAY18095.
 XX
 PT New Sel-1L nucleic acid molecule useful in the treatment of
 PT Alzheimer's disease, diabetes and cancer
 CC
 PS Claim 2; Page 71-73; 77pp; English.
 XX
 CC This sequence encodes a Sel-1L (Sel-1 like) protein (also previously
 CC known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and
 CC the host cell containing it can be used to prepare a Sel-1L protein.
 CC Compositions containing the Sel-1L proteins, or peptides that interfere
 CC with their binding can be used in a method for treating or preventing
 CC Alzheimer's disease, diabetes (especially insulin dependent diabetes
 CC mellitus), cancer (especially pancreatic cancer), stroke, vascular
 CC dementia, Parkinson's disease, or coronary heart disease. The
 CC compositions can also be used to treat conditions requiring modulation of
 CC fat or cholesterol metabolism.
 CC
 SO Sequence 3666 BP; 956 A; 881 C; 953 G; 874 T; 2 other;

Query Match 64.8%; Score 16.2; DB 20; Length 3666;
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcaggtgcacgcgcagcattt 21
 ||| ||| || ||||| |||||
 DB 1148 GCATGTGAATTGCCAGCATTT 1128

RESULT 36
 AAC77012/C
 ID AAC77012 standard; DNA: 3667 BP.
 XX
 AC AAC77012;

XX
 DT 10-AUG-1999 (first entry)
 XX
 DE Full length Mouse Sel-1L coding sequence.
 XX
 KW Sel-1L; Sel-1 like protein; Hip-1; Alzheimer's disease; diabetes; cancer;
 KW insulin dependent diabetes mellitus; pancreatic cancer; stroke; therapy;
 KW vascular dementia; Parkinson's disease; coronary heart disease;
 KW fat metabolism; cholesterol metabolism; ss.
 XX
 OS Mus sp.
 XX
 PN MO9927088-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 19-NOV-1998; 98WO-CA01058.
 XX
 PR 28-JUL-1998; 98US-0123549.
 PR 19-NOV-1997; 97US-0066140.
 XX
 PA (MOUN) MOUNT SINAI HOSPITAL.
 XX
 PI Bernstein A, Donoviel D;
 XX
 DR WPI: 1999-357833/30.
 DR P-PSDB; AAY18095.
 XX
 PT New Sel-1L nucleic acid molecule useful in the treatment of
 PT Alzheimer's disease, diabetes and cancer
 CC
 PS Claim 2; Page 66-69; 77pp; English.
 XX
 CC This sequence encodes a Sel-1L (Sel-1 like) protein (also previously
 CC known as Hip-1) of the invention. A vector containing the Sel-1L DNA, and
 CC the host cell containing it can be used to prepare a Sel-1L protein.
 CC Compositions containing the Sel-1L proteins, or peptides that interfere
 CC with their binding can be used in a method for treating or preventing
 CC Alzheimer's disease, diabetes (especially insulin dependent diabetes
 CC mellitus), cancer (especially pancreatic cancer), stroke, vascular
 CC dementia, Parkinson's disease, or coronary heart disease. The
 CC compositions can also be used to treat conditions requiring modulation of
 CC fat or cholesterol metabolism.
 CC
 SO Sequence 3667 BP; 956 A; 881 C; 954 G; 874 T; 2 other;

Query Match 64.8%; Score 16.2; DB 20; Length 3667;
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcaggtgcacgcgcagcattt 21
 ||| ||| || ||||| |||||
 DB 1148 GCATGTGAATTGCCAGCATTT 1128

RESULT 37
 AAC76014/C
 ID AAC76014 standard; CDNA: 5773 BP.
 XX
 AC AAC76014;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF1569 polynucleotide sequence SEQ ID NO:3137.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
 KW anticonvulsant; osteoplastic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;

XX	neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW	neurodegenerative disorder; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW	cholesterol ester storage; systemic lupus erythematosus; infection;
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW	bone damage; cartilage damage; antinflammatory disease; coagulation;
KW	thrombosis; contraceptive; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200058473-A2.
XX	
PD	05-OCT-2000.
XX	
PF	31-MAR-2000; 2000WO-US08621.
XX	
PR	31-MAR-1999; 99US-0127607.
PR	02-APR-1999; 99US-0127636.
PR	05-APR-1999; 99US-0127728.
PR	30-MAR-2000; 2000US-0540763.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shimkets RA, Leach M;
XX	
DR	WPI, 2000-602362/57.
DR	P-PsDB; AAB41805.
XX	
PT	Novel nucleic acids and peptides derived from open reading frame X,
PT	useful for treating e.g. cancers, proliferative disorders,
PT	neurodegenerative disorders and cardiovascular disease -
XX	
XX	Claim 5; Page 2354-2357; 5507pp; English.

AAC/4444 to AAC/7760 encode the proteins given AAB40237 to AAB43397,
which represent the human OREF open reading frames 1 to 3161. The OREF
sequences have activities such as: cytosolic; hepatotropic; vulnery;
antiproliferic; antiparkinsonian; nootropic; neuroprotective;
osteoplastic; anticoagulant; antiarthritic; immunosuppressant;
immunostimulant; cardiatic; fibroblastic; coagulant; vasotrophic;
antidiabetic; hypotensive; dermatological; immunosuppressive;
antitumour; antibacterial; antiviral; antifungal; antihemagic;
antihypid; and antianemic. The sequences can be used for determining
the presence of or predisposition to, or preventing or treating
pathological conditions associated with an OREF-associated disorder. The
nucleic acids can be used to express OREF proteins in gene therapy.
The proteins and nucleic acids may be used to treat cancers,
proliferative disorders, neurodegenerative disorders, osteoarthritis,
graft vs host disease, cardiovascular disease, diabetes mellitus,
hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
bacterial or fungal infection, malaria, autoimmune disorders, asthma,
allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
nocturnal haemoglobinuria, antinflammatory disease, to enhance
coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 5773 BP, 1691 A, 1125 C, 1269 G, 1687 T, 1 other;

Query Match	64.8%;	Score 16.2;	DB 21;	Length 5773;
Best Local Similarity	85.7%;	Pred. No. 2.4e+02;		
Matches 18; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0

```
QY      4  ggtgcatcgccagcalttcca 24
          || |||||
Db      4890 GGATCATCGCCAGCATTCAA 4870
```

RESULT	38
AAx76578/c	
ID	AAx76578 standard; cDNA; 7885 bp.
XX	
AC	AAx76578;

XX	10-AUG-1999	(first entry)	
DT			
XX			
DE	Human pancreas-specific tumour suppressor TSA305 gene SEQ ID NO:3.		
XX			
KW	Human; pancreas; TSA305; pancreatic cancer; carcinoma; diagnosis;		
KW	therapy; tumour suppressor; ss.		
XX			
OS	Homo sapiens.		
XX			
PH	Key	Location/Qualifiers	
FT	CDS	46..2430	
FT		/*lag- a	
XX			
PN	WO928457-A1.		
XX			
PD	10-JUN-1999.		
XX			
PF	25-NOV-1998;	98WO-IP05306.	
XX			
PR	20-APR-1998;	98JP-0126803.	
PR	28-NOV-1997;	97JP-0343789.	
XX			
PA	(SAKA) OTSUKA PHARM CO LTD.		
XX			
PL	Harada Y, Ozaki K;		
XX			
DR	WPI; 1999-358128/30.		
DR	P-PSDB; AAY17750.		
XX			
PT	Tumour suppressor gene TSA305 expressed specifically in pancreas,		
PT	useful for diagnosis and treatment of pancreatic cancer		
XX			
XX			
PS	Disclosure; Page 43-52; 54pp; Japanese.		
XX			
CC	The present sequence represents a gene designated TSA305, which is		
CC	expressed specifically in pancreas tissue. The gene, sequences		
CC	hybridizing with it, its expression product, and antibodies recognizing		
CC	the expression product, are useful in the investigation, diagnosis,		
CC	prevention and treatment of pancreatic cancer.		
XX			
XX	Sequence 7885 BP; 2332 A; 1466 C; 1619 G; 2468 T; 0 other;		

Query Match	64.8%;	Score 16.2;	DB 20;	Length 7885;
Best Local Similarity	85.7%;	Pred. No. 2.5e+02;		
Matches 18; Conservative	0;	Mismatches 3;	Indels 0;	Gaps 0;

```
QY      1  gcaggtgcacgccagcatt 21
          ||| ||| | ||| ||| |||
Db    1280 GCATGTGAATTGCCAGCATTT 1260
```

RESULT	39
AAC75460/c	
ID	AAC75460 standard; cDNA; 8028 BP

KM Human: crenrating frame: ORFXdetection: cyrostatic; hepatotropic;
KM vulnerrary: antiporiatic; antiparkinsonian; neuroptic; neuroproctive;
KM anticonvulsant; osteopthic; antiarrhythmic; immunosuppressant; cardiac;
KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KM hypotensive; dermatological; immunosuppressive; antinflammatory;
KM antiviral; antibacterial; antifungal; antirheumatic; antihypert;
KM antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
KM cardiovascular disease; diabetes mellitus; erythroidism; SCID; AIDS;
KM cholesterol ester storage; systemic lupus erythematosus; infection;

PF 08-OCT-1999; 99WO-US23573.
 XX
 PR 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 PA (CHTR) CHIRON CORP.
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 XX
 DR WPI: 2000-318079/27.

PT Isolated nucleotide sequences of *Neisseria meningitidis* which can be
 PT used in the diagnosis and treatment of *N. meningitidis* infection and
 PT other *Neisseria* infections, for example, *N. gonorrhoea* -
 XX
 PS Claim 7; Page 383-397; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic
 CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB256620 to AAB25663 represent
 CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to
 CC AAA81452 represent *Neisseria meningitidis* MemB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC *Neisseria* bacteria. For example, some of the identified proteins could
 CC be components of vaccines against *Meningococcus B*; against all serotypes;
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.
 XX

SQ Sequence 47475 BP; 11600 A; 13758 C; 12290 G; 9826 T; 1 other;

Query Match 64.8%; Score 16.2; DB 21; Length 47475;
 Best Local Similarity 85.7%; Pred. No. 3.3e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 gtagcgcgcagcattcgat 25
 ||||| ||||| ||||| ||||| |||||
 Db 16044 gtagcgcgcagcattcgat 16064

RESULT 42

AAF28542/c
 ID AAF28542 standard; DNA: 66986 BP.
 XX

AC AAF28542;
 XX

DT 04-APR-2001 (first entry)
 XX

DE Genomic fragment #29.
 XX

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW bronchopulmonary; endocarditis; meningitis; ss.
 XX

OS *Moraxella catarrhalis*.
 XX

PN WO200078968-A2.
 PN

XX
 PD 28-DEC-2000.
 XX
 PF 16-JUN-2000; 2000WO-US16649.
 XX
 PF 18-JUN-1999; 99US-0140121.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lagace RE, Patterson C, Berg KL;
 PI WPI: 2001-041427/05.
 DR
 DR
 XX

PT Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises *Moraxella catarrhalis* nucleic
 PT acids -
 XX

PS Claim 1; Page 247-263; 545pp; English.

XX The present invention relates to a *Moraxella catarrhalis* genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAF28514-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. *M. catarrhalis* (*Branhamella catarrhalis*) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. *M. catarrhalis* is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.
 XX

SQ Sequence 66986 BP; 18889 A; 13427 C; 15112 G; 19558 T; 0 other;

Query Match 64.8%; Score 16.2; DB 22; Length 66986;
 Best Local Similarity 85.7%; Pred. No. 3.5e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gtagcgcgcagcattcgat 24
 ||||| ||||| ||||| ||||| |||||
 Db 23807 GCTGCATCAGCAGCATTTTCA 23787

RESULT 43

AAA81479/c
 ID AAA81479 standard; DNA: 69936 BP.
 XX

AC AAA81479;
 XX

DT 04-DEC-2000 (first entry)
 XX

DE *N. meningitidis* partial DNA sequence gmm_27 SEQ ID NO:27.
 XX

XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW *Meningococcus B*; MemB; ds.
 XX

OS *Neisseria meningitidis*.
 XX

PN WO200022430-A2.
 XX

PD 20-APR-2000.
 XX

PF 08-OCT-1999; 99WO-US23573.
 XX

PR 09-OCT-1998; 98US-0103794.
 XX

PR 30-APR-1999; 99US-0132068.
 XX

PA (CHTR) CHIRON CORP.
 XX

PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Masignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 PI

PR 28-FEB-2000; 2000GB-0004695.

XX (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.

XX Pliza M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;

XX WPI; 2000-647603/62.

DR
XX
PT Neisseria meningitidis B full length genome sequence and open reading
frames are used to detect, treat and prevent Neisserial infections -

XX
PS Claim 7; Appendix A; 692pp; English.

XX
CC The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC Neisseria nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.

XX
SQ Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 349980;

Best Local Similarity 85.7%; Pred. No. 4,1e+02;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 gtgcacgcacgacatcgat 25
||||| |||||||||
Db 160514 gtgcacgcacgacatcgat 160534

Search completed: October 9, 2001, 11:38:55
Job time: 1850 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:39:43 ; Search time 221.13 Seconds
(without alignments)
21.403 Million cell updates/sec

Title: US-09-396-196F-6

Perfect score: 25

Sequence: 1 gcaggtgcacgcgcacattcgat 25

Scoring table: IDENTITY-NUC
Gapop 10.0, Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database: Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	17.6	70.4	3084	1	US-08-551-437-1
6	17.6	70.4	3084	3	US-09-004-225-1
7	17.6	70.4	3084	4	US-09-084-346-1
8	17.2	68.8	3186	1	US-08-761-258-8
9	17.2	68.8	3186	2	US-08-977-306-8
10	17.2	68.8	3186	2	US-08-483-376-1
11	16.6	66.4	65	3	US-08-833-167-19
12	16.6	66.4	522	3	US-08-833-167-40
13	16.6	66.4	531	3	US-08-833-167-36
14	16.6	66.4	20235	1	US-07-642-734C-3
15	16.6	66.4	20235	3	US-08-439-009A-3
16	16.2	64.8	8878	1	US-08-759-444-2
17	16.2	64.8	9880	1	US-08-680-897-1
18	16.2	64.8	750	3	US-09-167-717-3
19	16.2	64.8	2263	2	US-08-687-865A-1
20	16.2	64.8	2263	4	US-09-043-711-1
21	16.2	64.8	4203	4	US-08-866-757-1
22	16.2	64.8	4203	4	US-09-153-593-1
23	16.2	64.8	11236	1	US-07-853-913-1
24	16.2	64.8	30001	1	US-08-125-468-1
25	16.2	64.8	30001	2	US-08-474-933-1
26	15.6	62.4	386	1	US-08-620-467A-9
27	15.6	62.4	386	1	US-08-348-572-9

C 28	15.6	62.4	386	3	US-09-041-090B-9	Sequence 9, Appl
C 29	15.6	62.4	827	4	US-08-998-416-315	Sequence 315, App
C 30	15.6	62.4	1126	1	US-08-233-788A-48	Sequence 48, Appl
C 31	15.6	62.4	3172	6	5196511-1	Patent No. 5196511
C 32	15.6	62.4	12732	4	US-09-060-756-1	Sequence 1, Appl
C 33	15.4	61.6	480	6	5208144-34	Patent No. 5208144
C 34	15.4	61.6	644	4	US-08-998-416-155	Sequence 155, App
C 35	15.4	61.6	1542	4	US-09-385-028-16	Sequence 16, Appl
C 36	15.4	61.6	2266	2	US-08-724-394A-18	Sequence 18, Appl
C 37	15.4	61.6	3900	2	US-08-356-060A-42	Sequence 42, Appl
C 38	15.4	61.6	3900	4	US-08-460-900C-42	Sequence 42, Appl
C 39	15.4	61.6	4104	4	US-08-881-706-1	Sequence 1, Appl
C 40	15.4	61.6	4434	2	US-08-540-406-5	Sequence 5, Appl
C 41	15.4	61.6	4434	3	US-08-656-055-5	Sequence 5, Appl
C 42	15.4	61.6	4434	4	US-08-954-668-5	Sequence 5, Appl
C 43	15.4	61.6	4434	5	PCT-US95-13233-5	Sequence 5, Appl
C 44	15.4	61.6	11604	4	US-09-385-028-13	Sequence 13, Appl
C 45	15.4	61.6	15079	4	US-09-385-028-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-401-068-7
Sequence 7, Application US/08401068
Patent No. 5859335
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.308
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/351,970
FILING DATE: 08-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product="Biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pired. No. 0.0057;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcacgacattcgat 25
|||||
DB 84 GCAGGTGCATGCCAGCATTTCGAT 108

RESULT 2

US-08-846-338-7

Sequence 7, Application US/08846338

Patent No. 5869719

GENERAL INFORMATION:

APPLICANT: Patton, David

TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5869719artis Corporation

STREET: 520 White Plains Road, P.O. Box 2005

CITY: Tarrytown

STATE: NY

COUNTRY: USA

ZIP: 10591

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/846,338

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1041 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1038

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /product= "biotin synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pired. No. 0.0057;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcacgacattcgat 25
|||||
DB 84 GCAGGTGCATGCCAGCATTTCGAT 108

RESULT 3

US-08-411-768B-1

Sequence 1, Application US/08411768B

Patent No. 6083712

GENERAL INFORMATION:

APPLICANT: Olwen Birch

APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:

ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Wordperfect

SOFTWARE: Version 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/411,768B

FILING DATE: 31-March-95

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 3124/92

FILING DATE: 02-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 2134/93

FILING DATE: 15-JUL-1993

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5872 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Escherichia coli

STRAIN: DSM498

IMMEDIATE SOURCE:

CLONE: pBO30A-15/9

FEATURE:

NAME/KEY: CDS

LOCATION: 117..1157

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 117

OTHER INFORMATION: /product= "Biotin synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /gene= "b10B"

OTHER INFORMATION: /number= 1

FEATURE:

NAME/KEY: CDS

LOCATION: 2295..3050

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 2295

OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"

OTHER INFORMATION: /product= "protein"

OTHER INFORMATION: /gene= "bioC"

OTHER INFORMATION: /number= 3

FEATURE:

NAME/KEY: CDS

LOCATION: 3750..5039

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 3750

OTHER INFORMATION: /EC_number= 2.6.1.62

OTHER INFORMATION: /product= "DAPA synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /gene= "bioA"

OTHER INFORMATION: /number= 5

OTHER INFORMATION: /standard_name=

OTHER INFORMATION: "S-adenosyl-L-methionine:8-amino-7-oxononanoate

OTHER INFORMATION: amino:transf."

FEATURE:

NAME/KEY: CDS

LOCATION: 3750..5039

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 3750

OTHER INFORMATION: /EC_number= 2.6.1.62

OTHER INFORMATION: /product= "DAPA synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /gene= "bioA"

OTHER INFORMATION: /number= 5

OTHER INFORMATION: /standard_name=

OTHER INFORMATION: "S-adenosyl-L-methionine:8-amino-7-oxononanoate

OTHER INFORMATION: amino:transf."

FEATURE:

NAME/KEY: CDS

LOCATION: 3750..5039

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 3750

OTHER INFORMATION: /EC_number= 2.6.1.62

OTHER INFORMATION: /product= "DAPA synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /gene= "bioA"

OTHER INFORMATION: /number= 5

OTHER INFORMATION: /standard_name=

OTHER INFORMATION: "S-adenosyl-L-methionine:8-amino-7-oxononanoate

OTHER INFORMATION: amino:transf."

```
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter ptac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter ptac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioa RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent"
OTHER INFORMATION: /transcriptional_terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter ptac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match      100.0%; Score 25; DB 3; Length 5872;
Best local Similarity 100.0%; Pred. No. 0.0075;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcagtgcatcgccagcattcgat 25
    |||
Db 200 GCAGTGCATCGCCAGCATTTGAT 224

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pRO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biOf"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biOd"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "biOf RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biOd RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
```

FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0075;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcatgcgcagcatcgcattcgat 25
|||||
DB 200 GCAGGTGCATGCGCAGCATTCGCAT 224

RESULT 5
US-08-551-437-1
Sequence 1, Application US/08551437
Patent No. 5824545
GENERAL INFORMATION:
APPLICANT: Koltin, Yigal
APPLICANT: Riggle, Perry
APPLICANT: Gavrias, Vicki
APPLICANT: Bulawa, Chris
APPLICANT: Winter, Ben
TITLE OF INVENTION: IDENTIFICATION OF EUKARYOTIC
TITLE OF INVENTION: GROWTH-RELATED GENES AND PROMOTER ISOLATION VECTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/551,437
FILING DATE: 01-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 06286/009001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-507
TELEFAX: 617/542-890
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3084 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...3081
US-08-551-437-1

Query Match 70.4%; Score 17.6; DB 1; Length 3084;
Best Local Similarity 83.3%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 caggtgcatgcgcagcatcgcattcgat 25
|||||
DB 1691 CAGGTGCATTCGTCGATTCGCAT 1714

RESULT 6
US-09-004-225-1
Sequence 1, Application US/09004225
Patent No. 6020133
GENERAL INFORMATION:
APPLICANT: Koltin, Yigal
APPLICANT: Riggle, Perry
APPLICANT: Gavrias, Vicki
APPLICANT: Bulawa, Chris
APPLICANT: Winter, Ben
TITLE OF INVENTION: IDENTIFICATION OF EUKARYOTIC
TITLE OF INVENTION: GROWTH-RELATED GENES AND PROMOTER ISOLATION VECTOR
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,225
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/551,437
FILING DATE: 01-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 06286/009001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-507
TELEFAX: 617/542-890
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3084 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1...3081
US-09-004-225-1

Query Match 70.4%; Score 17.6; DB 3; Length 3084;
Best Local Similarity 83.3%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 caggtgcatgcgcagcatcgcattcgat 25
|||||
DB 1691 CAGGTGCATTCGTCGATTCGCAT 1714

RESULT 7
US-09-084-346-1
Sequence 1, Application US/09084346
Patent No. 6251593
GENERAL INFORMATION:
APPLICANT: Koltin, Yigal
APPLICANT: Riggle, Perry
APPLICANT: Gavrias, Vicki
APPLICANT: Bulawa, Chris
APPLICANT: Winter, Ben

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/977,306
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
TELEPHONE: (919) 541-8587
TELEFAX: (919) 541-8689
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 3186 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Pseudomonas fluorescens
STRAIN: CGA267356 (aka MCG134 and aka BL915)
IMMEDIATE SOURCE:
CLONE: pCIB146
FEATURE:
NAME/KEY: RBS
LOCATION: 245..251
OTHER INFORMATION: /note="potential ribosome binding
OTHER INFORMATION: site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 256..3006
OTHER INFORMATION: /product="Lema"
OTHER INFORMATION: /note="Lema coding sequence."
US-08-977-306-8

Query Match 68.8%; Score 17.2; DB 2; Length 3186;
Best Local Similarity 86.4%; Pred. No. 27;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 caggtgcatgcgcagcatctcg 23
||||| ||||| ||||| ||||| |||||

Db 2763 CAGGAGCATGCGCAGCATGTCG 2742

RESULT 10
US-08-483-376-1
Sequence 1, Application US/08483376
Patent No. 5955330
GENERAL INFORMATION:
APPLICANT: Vasil, Vmla
APPLICANT: Clancy, Maureen A.
APPLICANT: Ferl, Robert J.
APPLICANT: Vasil, Indra K.
APPLICANT: Hannah, L. C.
TITLE OF INVENTION: No. 5955330el Means for Enhancing Gene
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483,376
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/418,540
FILING DATE: 07-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/102,115
FILING DATE: 04-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/830,956
FILING DATE: 05-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/353,854
FILING DATE: 18-MAY-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 10-94B
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8080
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6386 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Zea mays
STRAIN: Black Sweet
FEATURE:
NAME/KEY: exon
LOCATION: 131..182
FEATURE:
NAME/KEY: exon
LOCATION: 1211..1324
FEATURE:
NAME/KEY: exon
LOCATION: 1828..1948
FEATURE:
NAME/KEY: exon
LOCATION: 2041..2187
FEATURE:
NAME/KEY: exon
LOCATION: 2269..2460
FEATURE:
NAME/KEY: exon
LOCATION: 2605..2728
FEATURE:
NAME/KEY: exon
LOCATION: 2822..3038
FEATURE:
NAME/KEY: exon
LOCATION: 3256..3351
FEATURE:
NAME/KEY: exon
LOCATION: 3447..3620
FEATURE:
NAME/KEY: exon
LOCATION: 3702..3818
FEATURE:
NAME/KEY: exon
LOCATION: 3912..4078
FEATURE:
NAME/KEY: exon
LOCATION: 4158..4381
FEATURE:
NAME/KEY: exon
LOCATION: 4517..4835

FEATURE:
NAME/KEY: exon
LOCATION: 4768..5212
FEATURE:
NAME/KEY: exon
LOCATION: 5372..5510
FEATURE:
NAME/KEY: exon
LOCATION: 5636..5917
US-08-483-376-1

Query Match 68.0%; Score 17; DB 2; Length 6386;
Best Local Similarity 80.0%; Pred. No. 38;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gcagtgatcgccagcatttcgat 25
DB 5691 GGAGAACCATCGCTGCATTTCGAT 5715

RESULT 11
US-08-833-167-19
Sequence 19, Application US/08833167
Patent No. 6100070
GENERAL INFORMATION:
APPLICANT: ZURFLUH, LINDA L
APPLICANT: MCHESTER, CHARLES A
APPLICANT: MCKEARN, JOHN P
APPLICANT: KLEIN, BARBARA K
APPLICANT: FENG, YIQUING
APPLICANT: BRAFORD-GOLDBERG, SARAH R
APPLICANT: LEE, STEPHEN C
TITLE OF INVENTION: G-CSF RECEPTOR AGONISTS
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESS:
ADDRESSEE: DENNIS A BENNETT, G.D. SEARLE & CO.,
ADDRESSEE: CORPORATE PATENT DEPT.,
STREET: P.O. BOX 5110
CITY: CHICAGO
STATE: ILLINOIS
COUNTRY: USA
ZIP: 60680
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/833,167
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US 96/15935
FILING DATE: 04-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/004,382
FILING DATE: 05-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENNETT, DENNIS A
REFERENCE/DOCKET NUMBER: 2907/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-694-5402
TELEFAX: 314-694-9095
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 65 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA (synthetic)"
US-08-833-167-19

Query Match 66.4%; Score 16.6; DB 3; Length 65;
Best Local Similarity 82.6%; Pred. No. 28;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 aggtgcatcgccagcatttcgat 25
DB 40 AGTGCAATGCCAGCATTTGCAT 62

RESULT 12
US-08-833-167-40
Sequence 40, Application US/08833167
Patent No. 6100070
GENERAL INFORMATION:
APPLICANT: ZURFLUH, LINDA L
APPLICANT: MCHESTER, CHARLES A
APPLICANT: MCKEARN, JOHN P
APPLICANT: KLEIN, BARBARA K
APPLICANT: FENG, YIQUING
APPLICANT: BRAFORD-GOLDBERG, SARAH R
APPLICANT: LEE, STEPHEN C
TITLE OF INVENTION: G-CSF RECEPTOR AGONISTS
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESS:
ADDRESSEE: DENNIS A BENNETT, G.D. SEARLE & CO.,
ADDRESSEE: CORPORATE PATENT DEPT.,
STREET: P.O. BOX 5110
CITY: CHICAGO
STATE: ILLINOIS
COUNTRY: USA
ZIP: 60680
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/833,167
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US 96/15935
FILING DATE: 04-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/004,382
FILING DATE: 05-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENNETT, DENNIS A
REFERENCE/DOCKET NUMBER: 2907/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-694-5402
TELEFAX: 314-694-9095
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 522 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA (synthetic)"
US-08-833-167-40

Query Match 66.4%; Score 16.6; DB 3; Length 522;
Best Local Similarity 82.6%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 aggtgcatcgccagcatttcgat 25
DB 33 AGTGCAATGCCAGCATTTGCAT 55

RESULT 13
US-08-833-167-36
Sequence 36, Application US/08833167
Patent No. 6100070
GENERAL INFORMATION:
APPLICANT: ZURELUH, LINDA L.
APPLICANT: MCMHURTER, CHARLES A
APPLICANT: MCKEARN, JOHN P
APPLICANT: KLEIN, BARBARA K
APPLICANT: FENG, YI QING
APPLICANT: BRAFORD-GOLDBERG, SARAH R
APPLICANT: LEE, STEPHEN C
TITLE OF INVENTION: G-CSF RECEPTOR AGONISTS
NUMBER OF SEQUENCES: 129
CORRESPONDENCE ADDRESS:
ADDRESSEE: DENNIS A BENNETT, G.D., SEARLE & CO.,
ADDRESSEE: CORPORATE PATENT DEPT.,
STREET: P.O. BOX 5110
CITY: CHICAGO
STATE: ILLINOIS
COUNTRY: USA
ZIP: 60680
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/833.167
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US 96/15935
FILING DATE: 04-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/004,382
FILING DATE: 05-OCT-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENNETT, DENNIS A
REFERENCE/DOCKET NUMBER: 2907/1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 314-694-5402
TELEFAX: 314-694-9095
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 531 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA (synthetic)"
US-08-833-167-36

Query Match 66.4%; Score 16.6; DB 3; Length 531;
Best Local Similarity 82.6%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 agtgcacgcgcacattcgat 25
||||| ||||||| ||
Db 33 AGGTGCATGCCAGCATTTGCAT 55

RESULT 14
US-07-642-734C-3/C
Sequence 3, Application US/07642734C
Patent No. 5824513
GENERAL INFORMATION:
APPLICANT: Katz, L
APPLICANT: Donadio, S
APPLICANT: McAlpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing

TITLE OF INVENTION: Erythromycin Analogs
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/Ab6D-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckerts, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722

```

OTHER INFORMATION: /function= "approximate span of
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene-eryA"
OTHER INFORMATION: /product= "Orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function= "approximate span of

```

```

OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-07-642-734C-3

```

```

Query Match 66.4%; Score 16.6; DB 1; Length 20235;
Best Local Similarity 82.6%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 gcagggtgcatgcgcagcatctcg 23
Db 4750 GCGGGTCATCGCCAGCGCTTCG 4728

```

```

RESULT 15
US-08-439-009A-3/C
; Sequence 3, Application US/08439009A
; Patent No. 6004787
GENERAL INFORMATION:
APPLICANT: Donadio, S
APPLICANT: Katz, L
APPLICANT: McAlpine, J B
TITLE OF INVENTION: Method of Directing Biosynthesis of
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven F. Weinstock
STREET: Abbott Laboratories D377/AbPD-2 One Abbott
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952-US-D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea

```

```

STRAIN: NRRL 238
FEATURE:
  NAME/KEY: CDS
  LOCATION: 19..10722
  OTHER INFORMATION: /codon_start= 19
  OTHER INFORMATION: /function= "gene eryA"
  OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
  OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 19..4470
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: module 3"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 97..1482
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 1693..2670
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 3406..3921
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 4171..4428
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 4471..10722
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: module 4"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 4471..5847
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 4"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 6054..7026
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 7165..9216
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 9433..9984
  OTHER INFORMATION: /function= "approximate span
  OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 10225..10483
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
  NAME/KEY: CDS
  LOCATION: 10723..20235
  OTHER INFORMATION: /codon_start= 10723
  OTHER INFORMATION: /function= "gene eryA"
  OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
  OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 10723..15165

```

```

OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 10831..12174
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 12379..13350
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 14062..14610
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 14857..15114
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 15166..20235
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: module 6"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 15172..16569
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 16768..17721
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 18379..18921
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 19149..19398
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
  NAME/KEY: misc_feature
  LOCATION: 19492..20235
  OTHER INFORMATION: /function= "approximate span of
  OTHER INFORMATION: chloesterase domain of module 6"
US-08-439-009A-3

Query Match          66.4%; Score 16.6; DB 3; Length 20235;
Best Local Similarity 82.6%; Pred. No. 70;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcaggtgcatgcgcagcatctg 23
   ||| ||| ||| ||| ||| |||
Db 4750 GCGGGTCATCGCCAGCGCTCG 4728

RESULT 16
US-08-759-444-2
; Sequence 2, Application US/08759444
; Patent No. 5824309
; GENERAL INFORMATION:
; APPLICANT: Dassarma, Shiladitya
; APPLICANT: Morshed, Fazela
; APPLICANT: Stuart, Elizabeth
; APPLICANT: Black, Samuel

```

TITLE OF INVENTION: RECOMBINANT GAS VESICLES AND USRS THEREOF
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/759,444
FILING DATE: 05-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/008,200
FILING DATE: 05-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Melkielejohn, Ph.D., Anita L
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07880/003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 8878 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-759-444-2

Query Match 64.8%; Score 16.2; DB 1; Length 8878;
Best Local Similarity 85.7%; Pred. No. 94;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 aggtgcacgcacgacattcg 23
||| ||||| ||||| |||
Db 1602 AGGAGCATCTCCAGCATCTCG 1622

RESULT 17
US-08-680-897-1
Sequence 1, Application US/08680897
Patent No. 6008051
GENERAL INFORMATION:
APPLICANT: Dassarma, Shiladitya
APPLICANT: Ng, Wei-lap
TITLE OF INVENTION: RECOMBINANT VECTOR AND PROCESS FOR CELL
TITLE OF INVENTION: FLUTATION
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: U.S.A.
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/680,897
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/271,270
FILING DATE:
APPLICATION NUMBER: 944,581
FILING DATE: 14-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Digilio, Frank S.
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8680
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9880 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-680-897-1

Query Match 64.8%; Score 16.2; DB 3; Length 9880;
Best Local Similarity 85.7%; Pred. No. 95;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 aggtgcacgcacgacattcg 23
||| ||||| ||||| |||
Db 2602 AGGAGCATCTCCAGCATCTCG 2622

RESULT 18
US-09-167-717-3/C
Sequence 3, Application US/09167717
Patent No. 613014
GENERAL INFORMATION:
APPLICANT: MUKOYAMA, MASAHARU
APPLICANT: YASUDA, SHINZO
APPLICANT: KOMATSUZAKI, SATOMI
TITLE OF INVENTION: MALEATE ISOMERASE GENE
FILE REFERENCE: PH-55505
CURRENT APPLICATION NUMBER: US/09/167,717
CURRENT FILING DATE: 1998-10-07
EARLIER APPLICATION NUMBER: JP 276261/1997
EARLIER FILING DATE: 1997-10-08
NUMBER OF SEQ ID NOS: 14
SEQ ID NO 3
LENGTH: 750
TYPE: DNA
ORGANISM: Pseudomonas fluorescens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(750)
US-09-167-717-3

Query Match 64.0%; Score 16; DB 3; Length 750;
Best Local Similarity 79.2%; Pred. No. 78;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 caagtcacgcacgacattcgat 25
||| ||||| ||| ||| |||
Db 81 CAGGAGCATCGCGGAGATCTCGGT 58

RESULT 19
US-08-687-865A-1/C
Sequence 1, Application US/08687865A
Patent No. 5955596
GENERAL INFORMATION:
APPLICANT: Jones, Kevin F.
APPLICANT: Zagursky, Robert J.

```

1  APPLICANT: Ooi, Peggy
2  TITLE OF INVENTION: The Nuca Protein of Haemophilus
3  TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein
4  NUMBER OF SEQUENCES: 23
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: American Cyanamid Company
7  STREET: One Cyanamid Plaza
8  CITY: Wayne
9  STATE: New Jersey
10 COUNTRY: U.S.A.
11 ZIP: 07470
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: Patentln Release #1.0, Version #1.30
18 CURRENT APPLICATION DATA:
19 APPLICATION NUMBER: US/08/687, 865A
20 FILING DATE: 26-JUL-1996
21 CLASSIFICATION: 536
22 ATTORNEY/AGENT INFORMATION:
23 NAME: Gordon, Alan M.
24 REGISTRATION NUMBER: 30,637
25 REFERENCE/DOCKET NUMBER: 33,250-00
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE: 201-831-3244
28 TELEFAX: 201-831-3305
29 INFORMATION FOR SEQ ID NO: 1:
30 SEQUENCE CHARACTERISTICS:
31 LENGTH: 2263 base pairs
32 TYPE: nucleic acid
33 STRANDEDNESS: single
34 TOPOLOGY: linear
35 MOLECULE TYPE: DNA (genomic)
36 FEATURE:
37 NAME/KEY: CDS
38 LOCATION: 229..2037
39 US-08-687-865A-1
40
41 Query Match 64.0%; Score 16; DB 2; Length 2263;
42 Best Local Similarity 79.2%; Pred. No. 93;
43 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0.
44
45 Oy 1 gcaggtgcacgcacgacatttcca 24
46 |||||11111111111111111111
47
48 Db 1541 GCGCGTACACCGCAGCATTTTGA 1518
49
50 RESULT 20
51 US-09-043-711-1/c
52 Sequence 1, Application US/09043711
53 Patent No. 6221365
54 GENERAL INFORMATION:
55 APPLICANT: Jones, Kevin F.
56 APPLICANT: Zagursky, Robert J.
57 APPLICANT: Ooi, Peggy
58 TITLE OF INVENTION: The Nuca Protein of Haemophilus
59 TITLE OF INVENTION: Influenzae and the Gene Encoding That Protein
60 NUMBER OF SEQUENCES: 23
61 CORRESPONDENCE ADDRESS:
62 ADDRESSEE: American Cyanamid Company
63 STREET: One Cyanamid Plaza
64 CITY: Wayne
65 STATE: New Jersey
66 COUNTRY: U.S.A.
67 ZIP: 07470
68
69 COMPUTER READABLE FORM:
70 MEDIUM TYPE: Floppy disk
71 COMPUTER: IBM PC compatible
72 OPERATING SYSTEM: PC-DOS/MS-DOS
73 SOFTWARE: Patentln Release #1.0, Version #1.30
74 CURRENT APPLICATION DATA:

```

```

1 APPLICATION NUMBER: US/09/043,711
2 FILING DATE:
3 CLASSIFICATION:
4 PRIOR APPLICATION DATA:
5 APPLICATION NUMBER: US 08/687,865
6 FILING DATE: 26-JUL-1996
7 ATTORNEY/AGENT INFORMATION:
8 NAME: Gordon, Alan M.
9 REGISTRATION NUMBER: 30,637
10 REFERENCE/DOCKET NUMBER: 33,250-00
11 TELECOMMUNICATION INFORMATION:
12 TELEPHONE: 201-831-3244
13 TELEFAX: 201-831-3305
14 INFORMATION FOR SEQ ID NO: 1:
15 SEQUENCE CHARACTERISTICS:
16 LENGTH: 2263 base pairs
17 TYPE: nucleic acid
18 STRANDEDNESS: single
19 TOPOLOGY: linear
20 MOLECULE TYPE: DNA (genomic)
21 FEATURE:
22 NAME/KEY: CDS
23 LOCATION: 229..2037
24 US-09-043-711-1
25
26 Query Match 64.0%; Score 16; DB 4; Length 2263;
27 Best Local Similarity 79.2%; Pred. No. 93;
28 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
29
30 QY 1 gcagtgcatcgccagcattcga 24
31 |||||
32 Db 1541 GCGCGTAGACCGCCAGCATTTTGA 1518
33
34 RESULT 21
35 US-08-866-757-1
36 Sequence 1, Application US/08866757
37 Patent No. 5858716
38 GENERAL INFORMATION:
39 APPLICANT: ELSHOURBAGY, NABIL A
40 APPLICANT: LI, XIAOTONG
41 APPLICANT: BERGSM, DERK J
42 TITLE OF INVENTION: NOVEL 7TM RECEPTOR (H2CAA71)
43 NUMBER OF SEQUENCES: 4
44 CORRESPONDENCE ADDRESS:
45 ADDRESSEE: RATNER & PRESTIA
46 STREET: P.O. BOX 980
47 CITY: VALLEY Forge
48 STATE: PA
49 COUNTRY: USA
50 ZIP: 19482
51 COMPUTER READABLE FORM:
52 MEDIUM TYPE: Diskette
53 COMPUTER: IBM Compatible
54 OPERATING SYSTEM: DOS
55 SOFTWARE: FastSeq for Windows Version 2.0
56 CURRENT APPLICATION DATA:
57 APPLICATION NUMBER: US/08/866,757
58 FILING DATE: 30-May-1997
59 CLASSIFICATION: 536
60 PRIOR APPLICATION DATA:
61 APPLICATION NUMBER:
62 FILING DATE:
63 ATTORNEY/AGENT INFORMATION:
64 NAME: PRESTIA, PAUL F
65 REGISTRATION NUMBER: 23,031
66 REFERENCE/DOCKET NUMBER: GH-70055
67 TELECOMMUNICATION INFORMATION:
68 TELEPHONE: 610-407-0700
69 TELEFAX: 610-407-0701
70 TELEX: 846169
71 INFORMATION FOR SEQ ID NO: 1:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 4203 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-866-757-1

Query Match 64.0%; Score 16; DB 2; Length 4203;
Best Local Similarity 79.2%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 caggtgcatcgccagcatctgcag 25
||||| | ||||| | ||||| | ||||| |
Db 1655 CAGGTGAACGCCATCATTTAGAT 1678

RESULT 22
US-09-153-593-1
Sequence 1, Application US/09153593A
Patent No. 6174994
GENERAL INFORMATION:
APPLICANT: ELSHOURBAGY, NABIL A
APPLICANT: LI, XIAOTONG
APPLICANT: BERGSMAN, DEBK J
TITLE OF INVENTION: NOVEL 7TM RECEPTOR (H2CA71)
FILE REFERENCE: GH-70055-1
CURRENT APPLICATION NUMBER: US/09/153,593A
CURRENT FILING DATE: 1998-09-15
EARLIER APPLICATION NUMBER: 08/866,757
EARLIER FILING DATE: 1997-05-30
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 4203
TYPE: DNA
ORGANISM: HOMO SAPIENS
US-09-153-593-1

Query Match 64.0%; Score 16; DB 4; Length 4203;
Best Local Similarity 79.2%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 caggtgcatcgccagcatctgcag 25
||||| | ||||| | ||||| | ||||| |
Db 1655 caggtgaacgccatcatctgcag 1678

RESULT 23
US-07-853-913-1
Sequence 1, Application US/07853913
Patent No. 5338839
GENERAL INFORMATION:
APPLICANT: McKay, Ronald D.G.
APPLICANT: Lendahl, Urban
TITLE OF INVENTION: Nestin Expression As An Indicator Of
TITLE OF INVENTION: Neuroepithelial Tumors
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/853,913
FILING DATE: 19920319
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/660,412
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,803
FILING DATE: 25-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/201,762
FILING DATE: 02-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4641AAAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 11236 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-07-853-913-1

Query Match 64.0%; Score 16; DB 1; Length 11236;
Best Local Similarity 79.2%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcaagtgcatcgccagcatctgcag 24
||||| | ||||| | ||||| | ||||| |
Db 1819 GCAGGAGCATCTCTGACATTTGCA 1842

RESULT 24
US-08-125-468-1
Sequence 1, Application US/08125468
Patent No. 5589385
GENERAL INFORMATION:
APPLICANT: Ryan, Michael J.
APPLICANT: Lotvin, Jason A.
APPLICANT: Strachy, Nancy
APPLICANT: Fantini, Susan E.
TITLE OF INVENTION: Cloning of the biosynthetic pathway for
TITLE OF INVENTION: chlorotetracycline and tetracycline formation and cosmids
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: USA
ZIP: 07470
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/125,468
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Tsevdos, Estelle J
REGISTRATION NUMBER: 31,145
REFERENCE/DOCKET NUMBER: 31,255-02

TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3241
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30001 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-125-468-1

Query Match 64.0%; Score 16; DB 1; Length 30001;
Best Local Similarity 79.2%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcaggtgcacgcagcattcga 24
||||| ||| ||||| |||||
Db 28819 GCAGGACGACGCCGACGATCGA 28842

RESULT 25
US-08-474-933-1
Sequence 1, Application US/08474933
Patent No. 5866410
GENERAL INFORMATION:
APPLICANT: Ryan, Michael J.
APPLICANT: Lotvin, Jason A.
APPLICANT: Strathy, Nancy
APPLICANT: Fantini, Susan E.
TITLE OF INVENTION: Cloning of the biosynthetic pathway for
TITLE OF INVENTION: chlortetracycline and tetracycline formation and cosmid
TITLE OF INVENTION: useful therein
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey
COUNTRY: USA
ZIP: 07470
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,933
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/125,468
FILING DATE: 22-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Tseydos, Estelle J
REGISTRATION NUMBER: 31,145
REFERENCE/DOCKET NUMBER: 31,255-02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201)831-3241
TELEFAX: (201)831-3305
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30001 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-474-933-1

Query Match 64.0%; Score 16; DB 2; Length 30001;
Best Local Similarity 79.2%; Pred. No. 1.4e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 gcaggtgcacgcagcattcga 24
||||| ||| ||||| |||||
Db 28819 GCAGGACGACGCCGACGATCGA 28842

RESULT 26
US-08-620-467A-9/C
Sequence 9, Application US/08620467A
Patent No. 5798231
GENERAL INFORMATION:
APPLICANT: HERCEND, THIERRY; TRIEBEL,
APPLICANT: FREDERIC; ROMAN-ROMAN, SERGIO; FERRADINI,
APPLICANT: LAURENT
TITLE OF INVENTION: NUCLEOTIDE SEQUENCE CODING
TITLE OF INVENTION: FOR VARIABLE REGIONS OF THE ' CHAINS OF HUMAN T
TITLE OF INVENTION: LYMPHOCYTE RECEPTORS, CORRESPONDING PEPTIDE
TITLE OF INVENTION: SEGMENTS AND DIAGNOSTIC AND THERAPEUTIC USES
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN, MUSERLIAN AND LUCAS,
ADDRESS: LLP
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/620,467A
FILING DATE: 22-MAR-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/348,572
FILING DATE: 19-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/934,529
FILING DATE: 24-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR92/00111
FILING DATE: 07-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 146.1159-CON-DIV-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-661-8000
TELEFAX: 212-661-8002
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 386
TYPE: NUCLEOTIDE
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: HUMAN
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE: HUMAN T LYMPHOCYTE
ORGANELLE:
FEATURE:
NAME/KEY: IGR a 11
LOCATION:

IDENTIFICATION METHOD:
OTHER INFORMATION: SEQUENCE V 7
US-08-620-467A-9

Query Match 62.4%; Score 15.6; DB 1; Length 386;
Best Local Similarity 81.8%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 ggtcgcgcgcagcattcgat 25
||||| ||||| ||||| |||||
Db 247 GGTGCTTCCGACGACATGTGCT 226

RESULT 27
US-08-348-572-9/c
Sequence 9, Application US/08348572
Patent No. 5817511
GENERAL INFORMATION:
APPLICANT: HERCEND, THIERRY; TRIEBEL, FREDERIC;
APPLICANT: ROMAN-ROMAN, SERGIO; FERRADINI, LAURENT
TITLE OF INVENTION: NUCLEOTIDE SEQUENCE CODING FOR VARIABLE
TITLE OF INVENTION: REGIONS OF THE ALPHA CHAINS OF HUMAN T LYMPHOCYTE RECEPTORS,
TITLE OF INVENTION: CORRESPONDING PEPTIDE SEGMENTS AND THE DIAGNOSTIC AND
TITLE OF INVENTION: THERAPEUTIC USES
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/348,572
FILING DATE: 19-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/934,529
FILING DATE: 24-NOV-1992
APPLICATION NUMBER: PCT/FR 92/00111
FILING DATE: 07-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 146.1159
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-661-8000
TELEFAX: 212-661-8002
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 386
TYPE: NUCLEOTIDE
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: HUMAN
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE: HUMAN T LYMPHOCYTE
ORGANELLE:
FEATURE:
NAME/KEY: IGR a 11

LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: SEQUENCE V Alpha 7
US-08-348-572-9

Query Match 62.4%; Score 15.6; DB 1; Length 386;
Best Local Similarity 81.8%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 4 ggtcgcgcgcagcattcgat 25
||||| ||||| ||||| |||||
Db 247 GGTGCTTCCGACGACATGTGCT 226

RESULT 28
US-09-041-090B-9/c
Sequence 9, Application US/09041090B
Patent No. 6114516
GENERAL INFORMATION:
APPLICANT: HERCEND, THIERRY; TRIEBEL, FREDERIC;
APPLICANT: ROMAN-ROMAN, SERGIO; FERRADINI, LAURENT
TITLE OF INVENTION: NUCLEOTIDE SEQUENCE CODING FOR VARIABLE
TITLE OF INVENTION: REGIONS OF THE ALPHA CHAINS OF HUMAN T LYMPHOCYTE RECEPTORS,
TITLE OF INVENTION: CORRESPONDING PEPTIDE SEGMENTS AND THE DIAGNOSTIC AND
TITLE OF INVENTION: THERAPEUTIC USES
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIERMAN & MUSERLIAN
STREET: 600 THIRD AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,090B
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,572
FILING DATE: 19-APR-1995
APPLICATION NUMBER: PCT/FR 92/00111
FILING DATE: 07-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: CHARLES A. MUSERLIAN
REGISTRATION NUMBER: 19,683
REFERENCE/DOCKET NUMBER: 146.1159
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-661-8000
TELEFAX: 212-661-8002
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 386
TYPE: NUCLEOTIDE
STRANDEDNESS: DOUBLE
TOPOLOGY: LINEAR
MOLECULE TYPE: CDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: HUMAN
STRAIN:
INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:
HAPLOTYPE:
TISSUE TYPE:
CELL TYPE:
CELL LINE: HUMAN T LYMPHOCYTE
ORGANELLE:
FEATURE:

NAME/KEY: IGR a 11
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: SEQUENCE V Alpha 7
US-09-041-090B-9

Query Match 62.4%; Score 15.6; DB 3; Length 386;
Best Local Similarity 81.8%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 ggtgcatgccagcatttcg 25
DB 247 GGTGCTGCCAGCATTTGCT 226

RESULT 29
US-08-998-416-315/c
Sequence 315, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYPTI
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 315:
SEQUENCE CHARACTERISTICS:
LENGTH: 827 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1255RP
US-08-998-416-315

Query Match 62.4%; Score 15.6; DB 4; Length 827;
Best Local Similarity 81.8%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 caagtgatgccagcatttcg 23
DB 250 CAGGCCATGCCAGCATTTG 229

RESULT 30
US-08-233-788A-48/c
Sequence 48, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 1126 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(3..23, 27..944, 948..1124)
US-08-233-788A-48

Query Match 62.4%; Score 15.6; DB 1; Length 1126;
Best Local Similarity 81.8%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcagtgatgccagcatttc 22
DB 789 GCAGTTCTGCCAGCATTTG 768

RESULT 31
5196511-1/c
Patent No. 5196511
APPLICANT: PLOW, EDWARD F./D/SOUZA, STANLEY E.
GINSBERG, MARK H.
TITLE OF INVENTION: PEPTIDES AND ANTIBODIES THAT INHIBIT
INTEGRIN-LIGAND BINDING
NUMBER OF SEQUENCES: 31
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/444,777

FILING DATE: 01-DEC-1989
SEQ ID NO:1:
LENGTH: 3172
5196511-1

Query Match 62.4%; Score 15.6; DB 6; Length 3172;
Best Local Similarity 81.8%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcagtgatcgccagcattcg 22
|||||
DB 2054 GCAGTGCACCCGCGCTCTGC 2033

RESULT 32
US-09-060-756-1
Sequence 1, Application US/09060756
Patent No. 6183957
GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO: 1
LENGTH: 12732
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-09-060-756-1

Query Match 62.4%; Score 15.6; DB 4; Length 12732;
Best Local Similarity 81.8%; Pred. No. 1.9e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 ggtgcatcgccagcattcgat 25
|||||
DB 11638 gctgcagcgccagcattcgat 11659

RESULT 33
5208144-34
Patent No. 5208144
APPLICANT: SMITH, JOHN A.; RAYCHOMHURY, RAKTIMA; NILES, JOHN L.
TITLE OF INVENTION: METHOD FOR DETECTION OF HUMAN DNA
CONTAINING THE GENE ENCODING LOW DENSITY LIPOPROTEIN RECEPTOR
NUMBER OF SEQUENCES: 42
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/396,697
FILING DATE: 22-AUG-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 313,682
FILING DATE: 22-FEB-1989
APPLICATION NUMBER: 235,211
FILING DATE: 23-AUG-1988
SEQ ID NO:34:
LENGTH: 480
5208144-34

Query Match 61.6%; Score 15.4; DB 6; Length 480;
Best Local Similarity 76.0%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcagtgatcgccagcattcgat 25

DB 295 gcagtgatcgccagcattggat 319
|||||

RESULT 34
US-08-998-416-155/c
Sequence 155, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Redischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 155:
SEQUENCE CHARACTERISTICS:
LENGTH: 644 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE: PAG1059RP
ORGANISM: PAG1059RP
US-08-998-416-155

Query Match 61.6%; Score 15.4; DB 4; Length 644;
Best Local Similarity 76.0%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcagtgatcgccagcattcgat 25
|||||
DB 541 GCAAGTGCATCCACACTTTCCAT 517

RESULT 35
US-09-385-028-16
Sequence 16, Application US/09385028
Patent No. 6232106
GENERAL INFORMATION:
APPLICANT: Kusan E. Jensen
APPLICANT: Kwamena A Aidoo

```

; APPLICANT: Ashish S. Paradkar
; TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
; Patent No. 6232106
; TITLE OF INVENTION: Acid Biosynthesis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: JACOBSON, PRICE, HOLMAN & STERN, PLLC
; STREET: The Jenner Building, 400 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/385,028
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/790,462
; FILING DATE: 29-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: D. Douglas Price
; REGISTRATION NUMBER: 24,514
; REFERENCE/DOCKET NUMBER: 1418/P574520S2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 638-6666
; TELEFAX: (202) 39305350
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-385-028-16

Query Match          61.6%; Score 15.4; DB 4; Length 1542;
Best Local Similarity 76.0%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcagcattcgat 25
    |||||  ||  |||  |||||
Db 1425 GCAGGTGGTGGCGCGAGCTTTCGAT 1449

RESULT 36
US-08-724-394A-18
; Sequence 18, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
; TITLE OF INVENTION: Sequences and Antibodies Thereeto
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESS: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2266 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..2266
; OTHER INFORMATION: /note="cDNA 22B"
; US-08-724-394A-18

Query Match          61.6%; Score 15.4; DB 2; Length 2266;
Best Local Similarity 76.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcagcattcgat 25
    |||||  ||  |||||  |||
Db 971 GCAGGATCAGGTCAGCATTTTCGAT 995

RESULT 37
US-08-356-060A-42/C
; Sequence 42, Application US/08356060A
; Patent No. 5844079
; GENERAL INFORMATION:
; APPLICANT: Ingham, Phillip W.
; APPLICANT: McMahon, Andrew P.
; APPLICANT: Tablin, Clifford J.
; TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
; TITLE OF INVENTION: Proteins and Uses Related Thereeto
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESS: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,060A
; FILING DATE: 14-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/176,427
; FILING DATE: 30-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: HMT-006CP

```

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 3900 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3897
US-08-356-060A-42

Query Match
Best Local Similarity 61.6%; Score 15.4; DB 2; Length 3900;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 gcaggtgcacgcgcagcattcgat 25
Db 3492 GCAGGTCCTCCGCGAGGATCCGCAT 3468

RESULT 38
US-08-460-900C-42/C
Sequence 42, Application US/08460900C
Patent No. 6165747
GENERAL INFORMATION:
APPLICANT: Ingham, Phillip W.
APPLICANT: McMahon, Andrew P.
APPLICANT: Tadin, Clifford J.
APPLICANT: Bumcrot, David A.
APPLICANT: Marti-Gorostiza, Elisa
TITLE OF INVENTION: Vertebrate Embryonic Pattern-Inducing
TITLE OF INVENTION: Proteins and Uses Related Thereto
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY, HONG & ELIOT LLP
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,900C
FILING DATE: 5-JUNE-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/435,093
FILING DATE: 4-MAY-1995
APPLICATION NUMBER: US 08/356,060
FILING DATE: 14-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/176,427
FILING DATE: 30-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36,709
REFERENCE/DOCKET NUMBER: HMW-006.05
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 832-1000
TELEFAX: (617) 832-7000
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 3900 base pairs
TYPE: nucleic acid
```

```
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3897
US-08-460-900C-42

Query Match
Best Local Similarity 61.6%; Score 15.4; DB 4; Length 3900;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 gcaggtgcacgcgcagcattcgat 25
Db 3492 GCAGGTCCTCCGCGAGGATCCGCAT 3468

RESULT 39
US-08-881-706-1/C
Sequence 1, Application US/08881706
Patent No. 6245969
GENERAL INFORMATION:
APPLICANT: Chong, Joane
APPLICANT: Li, Jianming
TITLE OF INVENTION: Receptor Kinase BIN1
FILE REFERENCE: 07251/022001
CURRENT APPLICATION NUMBER: US/08/881,706
CURRENT FILING DATE: 1997-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 4104
TYPE: DNA
ORGANISM: Arabidopsis sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (97)..(3687)
US-08-881-706-1

Query Match
Best Local Similarity 61.6%; Score 15.4; DB 4; Length 4104;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 gcaggtgcacgcgcagcattcgat 25
Db 2624 GCAGTTCATCGCCAGAGTTCCAT 2600

RESULT 40
US-08-540-406-5/C
Sequence 5, Application US/08540406
Patent No. 5837538
GENERAL INFORMATION:
APPLICANT: SCOTT, MATHEW P.
APPLICANT: GOODRICH, LISA V.
APPLICANT: JOHNSON, RONALD L.
TITLE OF INVENTION: Patched Genes and their Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/540,406
FILING DATE: 06-Oct-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: a60190-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4434 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-540-406-5

Query Match 61.6%; Score 15.4; DB 2; Length 4434;
Best Local Similarity 76.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcagtgatcgccagcattcgat 25
|||||
DB 3654 GCAGCTGCTCGCGAGATCGCGAT 3630

RESULT 41
US-08-656-055-5/c
Sequence 5, Application US/08656055
Patent No. 6027882
GENERAL INFORMATION:
APPLICANT: SCOTT, MATHEW P
APPLICANT: GOODRICH, LISA V
APPLICANT: JOHNSON, RONALD L
TITLE OF INVENTION: Patched Genes and their Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/656,055
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/540,406
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertram I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: a60190-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4434 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-656-055-5

Query Match 61.6%; Score 15.4; DB 3; Length 4434;
Best Local Similarity 76.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcagtgatcgccagcattcgat 25
|||||
DB 3654 GCAGCTGCTCGCGAGATCGCGAT 3630

RESULT 42
US-08-954-668-5/c
Sequence 5, Application US/08954668
Patent No. 6172200
GENERAL INFORMATION:
APPLICANT: SCOTT, MATHEW P
APPLICANT: GOODRICH, LISA V
APPLICANT: JOHNSON, RONALD L
TITLE OF INVENTION: Patched Genes and their Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley, Hoag & Elliot
STREET: One Post Office Square
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII(text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/954,668
FILING DATE: 20-Oct-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Vincent, Matthew P.
REGISTRATION NUMBER: 36709
REFERENCE/DOCKET NUMBER: SUV-003.06
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-832-1000
TELEFAX: 617-832-7000
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4434 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-954-668-5

Query Match 61.6%; Score 15.4; DB 4; Length 4434;
Best Local Similarity 76.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcagtgatcgccagcattcgat 25
|||||
DB 3654 GCAGCTGCTCGCGAGATCGCGAT 3630

RESULT 43
PCT-US95-13233-5/c
Sequence 5, Application PC/TUS9513233
GENERAL INFORMATION:
APPLICANT: THE BOARD OF TRUSTEES OF TEH LELAND STANFORD JUNIOR UNIVERSITY
TITLE OF INVENTION: Patched Genes and their Use
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco
STATE: CA
COUNTRY: US
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13233
FILING DATE: 06-OCT-1990
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Rowland, Bertam I
REGISTRATION NUMBER: 20015
REFERENCE/DOCKET NUMBER: A60190-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4434 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US95-13233-5

Query Match 61.6%; Score 15.4; DB 5; Length 4434;
Best Local Similarity 76.0%; Pred. No. 2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcaggtgcatcgccagatcgcgat 25
||||| 111 111 11 11111
Db 3654 GCAGGTGCTCGCGAGGATGCGCAT 3630

RESULT 44
US-09-385-028-13
Sequence 13, Application US/09385028
Patent No. 6232106
GENERAL INFORMATION:
APPLICANT: Susan E. Jensen
APPLICANT: Kwamena A Aidoo
APPLICANT: Ashish S. Paradkar
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
Patent No. 6232106
TITLE OF INVENTION: Acid Biosynthesis
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
STREET: The Jennifer Building, 400 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,028
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/790,462
FILING DATE: 29-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: D. Douglas Price
REGISTRATION NUMBER: 24,514

REFERENCE/DOCKET NUMBER: 1418/P574520S2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 39305350
TELEX: RCA 248593 IDEA UR
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 11604 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-385-028-13

Query Match 61.6%; Score 15.4; DB 4; Length 11604;
Best Local Similarity 76.0%; Pred. No. 2.3e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcaggtgcatcgccagatcgcgat 25
||||| 111 111 1111111
Db 3332 GCAGGTGCTCGCGAGGATGCGCAT 3356

RESULT 45
US-09-385-028-1
Sequence 1, Application US/09385028
Patent No. 6232106
GENERAL INFORMATION:
APPLICANT: Susan E. Jensen
APPLICANT: Kwamena A Aidoo
APPLICANT: Ashish S. Paradkar
TITLE OF INVENTION: DNA Sequence Encoding Enzymes of Clavulanic
Patent No. 6232106
TITLE OF INVENTION: Acid Biosynthesis
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC
STREET: The Jennifer Building, 400 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/385,028
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/790,462
FILING DATE: 29-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: D. Douglas Price
REGISTRATION NUMBER: 24,514
REFERENCE/DOCKET NUMBER: 1418/P574520S2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 39305350
TELEX: RCA 248593 IDEA UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 15079 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Streptomyces clavuligerus

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19.2	76.8	538	30	AV439422 AV439422
2	19.2	76.8	548	110	AV748431 AV748431
3	19.2	76.8	715	256	B07718 73611E03703
4	19.2	76.8	801	226	AO288588 nbxb0033k
5	18.6	74.4	376	13	AA879064 o18a12.s
6	18.6	74.4	627	153	BG455205 NF102G09P
7	18.2	72.8	231	18	AI329089 a9912ne.r
8	18.2	72.8	403	234	AO849360 LMAJFV1.1
9	18.2	72.8	527	258	AL390686 Leishman1
10	18.2	72.8	834	239	AL2176195 SP_0141_A
11	17.8	71.2	360	258	AL390537 Leishman1
12	17.8	71.2	384	26	AV186249 AV186249
13	17.8	71.2	439	232	AO686803 nbxb00731
14	17.8	71.2	498	21	AI532390 SD03891.5
15	17.8	71.2	580	226	AO291570 nbxb0039K
16	17.8	71.2	693	145	BF203823 60186810
17	17.8	71.2	888	219	AL065360 Drosophila
18	17.8	71.2	1090	222	CNS0585 CNS0585
19	17.8	71.2	1862	192	AK018497 Mus muscu
20	17.6	70.4	317	163	BE089771 QVQ, BT070
21	17.6	70.4	418	235	AO890634 HS_2204_A
22	17.6	70.4	450	229	AO516531 HS_5052_B
23	17.6	70.4	475	151	BF654469 278665 MA
24	17.6	70.4	506	175	BG266805 1000101G1
25	17.6	70.4	549	31	AV619677 AV619677
26	17.6	70.4	558	233	AO795581 nbxb0056A
27	17.6	70.4	582	23	AI665214 605008C11
28	17.6	70.4	615	148	BF425225 sus4h10.y
29	17.6	70.4	1032	221	CNS032YD AL262303 Tetradon
30	17.6	70.4	1039	220	CNS038BD AL236110 Tetradon
31	17.6	70.4	1077	172	BG027526 BG0275377
32	17.4	69.6	376	171	BF933544 CM1-NT027
33	17.4	69.6	527	225	AO171736 HS_3172_B
34	17.4	69.6	634	227	AO350006 RSC111-11
35	17.4	69.6	786	221	CNS038HM AL244481 Tetradon
36	17.2	68.8	147	164	BE156980 RC3-HT037
37	17.2	68.8	309	10	AA686065 EST109111
38	17.2	68.8	435	240	AA273781 RPT-23-1
39	17.2	68.8	495	146	BF291841 MHE2204_F
40	17.2	68.8	519	258	TA47A090 AL455103 T. brucei
41	17.2	68.8	544	258	TA329B04Q AL492386 T. brucei
42	17.2	68.8	555	143	BF008069 1698729 A
43	17.2	68.8	565	239	AZ212380 Sheared D
44	17.2	68.8	570	11	AA769145 ob51H03.s
45	17.2	68.8	646	143	BF008062 1664118 A

ALIGNMENTS

RESULT 1
 LOCUS AV439422 538 bp mRNA EST 23-AUG-2000
 DEFINITION AV439422 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
 PS050C10_r 5', mRNA sequence.
 ACCESSION AV439422 GI:8594647
 VERSION AV439422.1 GI:8594647
 KEYWORDS EST.
 SOURCE Porphyra yezoensis.
 ORGANISM Porphyra yezoensis
 Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
 Porphyra.
 REFERENCE 1 (bases 1 to 538)
 Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and
 Tabata, S.
 TITLE Generation of 10,154 expressed sequence tags from a leafy
 gametophyte of a marine red alga, Porphyra yezoensis

JOURNAL DNA Res. 7, 223-227 (2000)
 MEDLINE 20363100
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
 source location/Qualifiers
 1..538
 /organism="Porphyra yezoensis"
 /strain="TU-1"
 /db_xref="taxon:2788"
 /clone="PS050C10_r"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 114 a 158 c 170 g 96 t
 ORIGIN

Query Match Score 19.2; DB 30; Length 538;
 Best Local Similarity 87.5%; Pred. No. 1.1e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 caggtgcatgcgcagcattccgat 25
 |||||
 Db 385 CAGGTGCATGCCCTGCATTTCAT 362

RESULT 2
 LOCUS AV748431 548 bp mRNA EST 19-OCT-2000
 DEFINITION AV748431 NPC Homo sapiens cDNA clone NPCAOF03 5', mRNA sequence.
 ACCESSION AV748431 GI:10906279
 VERSION AV748431.1 GI:10906279
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 548)
 Song, H., Peng, Y., Gu, Y., Yang, Y., Gao, G., Xiao, H., Xu, X., Li, N., Gu
 Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Deng, L., Xu, S., Gu
 W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Ye, M., Zhang, Q.,
 Han, Z., Chen, Z., Hu, R. and Chen, J.
 Homo sapiens NPC library cDNA clones
 Unpublished (2000)
 CONTACT: Qinghua Zhang
 Shanghai Institute of Endocrinology, Rui-Jin Hospital
 197 Rui-Jin II Road, Shanghai 200025, P. R. China
 Tel: 86-21-64370045(ex.663332)
 Fax: 86-21-64743206
 Email: mshl@ems.sh.cn
 This clone is available at Shanghai Hematology Institute in
 Shanghai.

FEATURES
 source location/Qualifiers
 1..548
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NPCAOF03"
 /clone_lib="NPC"
 /tissue_type="pituitary"
 /dev_stage="Adult"
 /lab_host="SOLR"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 147 a 146 c 131 g 119 t 5 others
 ORIGIN

Query Match Score 19.2; DB 110; Length 548;

Best Local Similarity 87.5%; Pred. No. 11e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 caggtgcatgcgcagcattcgat 25
|||||
Db 168 CAGGGGATCCGACGATTTCTAT 191

RESULT 3

LOCUS B07718 715 bp DNA GSS 15-NOV-1997
DEFINITION 73611E0370318 Rhodobacter sphaeroides 2.4.1 genomic DNA library
(P44953), DNA sequence.

ACCESSION B07718.1 GI:2055510

KEYWORDS GSS.

SOURCE Rhodobacter sphaeroides.

ORGANISM Rhodobacter sphaeroides

REFERENCE 1 (bases 1 to 715)
Choudhary, M., Mackenzie, C., Nereng, K., Sodergren, E., Weinstein, G.M.,
and Kaplan, S.

TITLE Low-resolution sequencing of Rhodobacter sphaeroides 2.4.1T:
chromosome II is a true chromosome

JOURNAL Microbiology 143, 3085-3099 (1997)

COMMENT

Contact: Mackenzie, C.
Department of Microbiology and Molecular Genetics
6431 Fannin St., Houston, TX 77006, USA
Tel: (713) 794-1742
Fax: (713) 794-1782
Email: mackenzie@umt.tmc.edu
Seq primer: pBluescript SK (-) T7
Class: shotgun.

FEATURES

source

Location/Qualifiers

1..715

/organism="Rhodobacter sphaeroides"

/strain="2.4.1T"

/db_xref="taxon:1063"

/clone="73611E0370318"

/clone_lib="Rhodobacter sphaeroides 2.4.1 genomic DNA
library"

/lab_host="E. coli S17-1"

/note="Vector: pLA2917. Genomic DNA from Rhodobacter
sphaeroides was prepared and partially digested with
Sau3A1. Size selected (20kb) fragments were subcloned into
the BglII site of the cosmid vector pLA2917 (Allen, L. N.
and R. S. Hanson, 1985. Journal of Bacteriology,
161:955-962. The library was then ordered around
chromosome II of Rhodobacter sphaeroides (Choudhary et
al., 1994. Journal of Bacteriology, 176:7694-7702). The
cosmids were then digested with a variety of restriction
enzymes (BamHI, EcoRI, BglII, PstI, EcoRV, NotI and
DnaSI) and restriction fragments subcloned into the
respective multiple cloning site sites of pBluescript SK
(-). Note BglII fragments were subcloned into the
pBluescript BamHI site. DnaSI fragments were subcloned
into the EcoRV site. All subclones were then sequenced
into E. coli XL1Blue MRF. All fragments were then sequenced
and the sequences where possible were assembled using the
GCG program GELASSEMBLE."

BASE COUNT 97 a 238 c 255 g 107 t 18 others
ORIGIN

Query Match 76.8%; Score 19.2; DB 256; Length 715;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcaggtgcatgcgcagcattcgat 24

Db 42 CAGGTGATCGCGACGCTCTTCA 65
|||||

RESULT 4

LOCUS A0288588 801 bp DNA GSS 03-DEC-1998
DEFINITION nxb0033K04f CUGI Rice BAC library Oryza sativa genomic clone
nxb0033K04f, DNA sequence.

ACCESSION A0288588
VERSION A0288588.1 GI:3950202

KEYWORDS GSS.

SOURCE Oryza sativa.

ORGANISM Oryza sativa

REFERENCE 1 (bases 1 to 801)
Wing, R.A. and Dean, R.A.

TITLE A BAC End Sequencing Framework to Sequence the Rice Genome

JOURNAL Unpublished (1998)

COMMENT Contact: Wing, R.A.
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TATACGACTCACTATAGCG
Class: BAC ends
High quality sequence stop: 406.

FEATURES

source

Location/Qualifiers

1..801

/organism="Oryza sativa"

/strain="Japonica"

/cultivar="Nipponbare"

/db_xref="taxon:4530"

/clone="nxb0033K04f"

/clone_lib="CUGI Rice BAC library"

/tissue_type="leaf"

/lab_host="E. coli DH10B"

/note="Vector: pBluescript II. Site 1: HindIII. Site 2:
HindIII. Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9%. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."

BASE COUNT 213 a 214 c 162 g 212 t
ORIGIN

Query Match 76.8%; Score 19.2; DB 226; Length 801;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 caggtgcatgcgcagcattcgat 25
|||||
Db 462 CCGGTGATCGCGACGATGTCAT 485

RESULT 5
AA879064/c 376 bp mRNA EST 24-SEP-1998
LOCUS oJ18a12.s1 NCI-CGAP.Kid5 Homo sapiens cDNA clone IMAGE:1492510 3'
DEFINITION similar to TR:014437 Q14437 GOLGI APPARATUS GP25L2 PROTEIN. ; mRNA
sequence.
ACCESSION AA879064
VERSION AA879064.1 GI:2988029
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 376)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/CLND at:
www.bio.lnlnl.gov/bbrp/image/image.html
Insert length: 453 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 159.
FEATURES
Location/Qualifiers
1..376
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1492510"
/clone_lib="NCI-CGAP.Kid5"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACGTGAGAAATCGCGCGCGCAATATTTTTTTTTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 91 a 93 c 100 g 92 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 13; Length 376;
Best Local Similarity 84.0%; Pred. No. 2.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcaagtgatcgccagcatctgat 25
|||||
Db 197 GCAGATGATCATCCTCTTCGAT 173

RESULT 6
BG455205 627 bp mRNA EST 19-MAR-2001
LOCUS NF102G09PULF1070 Phosphate starved leaf Medicago truncatula cDNA
DEFINITION clone NF102G09PL 5', mRNA sequence.
ACCESSION BG455205
VERSION BG455205.1 GI:13378530
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE 1 (bases 1 to 627)
AUTHORS Liu, J., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores
'H.R., Imman, J.T., Weller, J.W., May, G.D. and Harrison, M.J.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula phosphate-starved leaf library
JOURNAL Unpublished (2000)
COMMENT Contact: Harrison MJ
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7325
Fax: 580 221 7380
Email: mj.harrison@noble.org
Insert length: 627 Std Error: 0.00
Plate: 102 row: 6 column: 09
Seq primer: TCACACGACGAAACAGCATATGAC.
FEATURES
Location/Qualifiers
1..627
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF102G09PL"
/clone_lib="Phosphate starved leaf"
/tissue_type="leaf"
/dev_stage="trifoliolate"
/note="Vector: Lambda Zap; At the trifoliolate stage, M.
truncatula plants were transplanted to phosphate-free sand
and grown for a further 30 days. During this 30 day
period, the plants were fertilized twice weekly with 1/2
Hoaglands solution containing only 20um potassium
phosphate. RNA was prepared from above ground tissues. "

BASE COUNT 147 a 141 c 167 g 172 t
ORIGIN

Query Match 74.4%; Score 18.6; DB 153; Length 627;
Best Local Similarity 84.0%; Pred. No. 2.2e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcaagtgatcgccagcatctgat 25
|||||
Db 375 GCAGCTGATCGCCAGCATGTTGAT 351

RESULT 7
A1329089 231 bp mRNA EST 28-DEC-1998
LOCUS A1329089/c
DEFINITION a9g12ne.f1 Neurospora crassa evening cDNA library Neurospora crassa
cDNA clone a9g12ne 3', mRNA sequence.
ACCESSION A1329089
VERSION A1329089.1 GI:4065648
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
REFERENCE 1 (bases 1 to 231)
AUTHORS Zhu, H., Lai, H., Kupfer, D., Dunlap, J.C. and Roe, B.A.
TITLE Two Neurospora crassa EST Databases
JOURNAL Unpublished (1998)
COMMENT Other ESTs: a9g12ne.f1
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Reverse Primer
High quality sequence stop: 179.
FEATURES
Location/Qualifiers

```

source
1. .231
/organism="Neurospora crassa"
/strain="Strain 30-7 (bd; A)"
/db_xref="taxon:5141"
/clone="ag912ne"
/clone_lib="Neurospora crassa evening cDNA library"
/tissue_type="tissue harvested following 22hr growth in
dark"
/notes="vector: pBluescript SK-; site: 1: XbaI; site: 2:
EcoRI; See: Bell-Pedersen, D., et al. PNAS 93:13096, 1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3'
end of cDNA cloned into EcoRI site of pBluescript"

BASE COUNT
50 a 71 c 39 g 71 t

Query Match
Best Local Similarity 72.8%; Score 18.2; DB 18; Length 231;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 aggtgcacgcgcacattcgat 25
|||||
Db 61 AGGGCAGTCGCCAGCATTCGAT 39

RESULT 8
LOCUS A0849360 403 bp DNA GSS 13-APR-2001
DEFINITION LMAJFV1_lm45g09.y1 Leishmania major FV1 random genomic library
Leishmania major genomic clone LMAJFV1_lm45g09 5' similar to
TR:015206 O15206 AXONEMAL DYNEIN HEAVY CHAIN;; DNA sequence.
ACCESSION A0849360
VERSION A0849360.1 GI:6054008
KEYWORDS GSS.
SOURCE Leishmania major.
ORGANISM Leishmania major
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE
1 (bases 1 to 403)
Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L.,
Kissinger, J., Roos, D.S., Maria, M., Hillier, L., Chinwalla, A.,
Blistain, A., Schmitt, A., Person, B., Theising, B., Ritzer, E., Ronko,
I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M.,
Harvey, N., McCann, R., Tsagaris, R., Williams, T., Jackson, Y.,
Bowers, Y., Swaller, T., Waterston, R., Wilson, R. and Beverley, S.M.
A survey of the Leishmania major Friedlin strain VI genome by
shotgun sequencing: a resource for DNA
Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
Contact: Akopyants, NS / Beverley, SM
WashU Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain VI genome by shotgun sequencing' and the Washington
University Genome Sequencing Center For information on obtaining
clone material please contact: Natalia S. Akopyants Ph.D.
(natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@borcim.wustl.edu)
Trace considered overall poor quality
Seq primer: -40RP from Gibco
Class: Shotgun
High quality sequence stop: 1.
Location/Qualifiers
1. .403
/organism="Leishmania major"
/strain="Friedlin strain VI"
/db_xref="taxon:5664"

FEATURES
source

Qy 1 gcaagtcacgcgcacattcgat 23
|||||
Db 282 GCAAGTCGCGCGCAGCATTCGAT 304

RESULT 10

```

LOCUS	A2176195/c	834 bp	DNA	GSS	30-AUG-2000
DEFINITION	SP.011.1 AL A06.77A Strongylocentrotus purpuratus, purple sea urchin 'sperm genomic BAC library Strongylocentrotus purpuratus genomic clone plate-141 Col-11 Row-A, DNA sequence.				
ACCESSION	A2176195				
VERSION	A2176195.1	GI:8348570			
KEYWORDS	GSS.				
SOURCE	Strongylocentrotus purpuratus.				
ORGANISM	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.				
REFERENCE	1 (bases 1 to 834)				
AUTHORS	Cameron,R.A., Mahairas,G., Rast,J.P., Martinez,P., Biondi,T.R., Swartell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H. and Hood,L.				
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources				
JOURNAL	Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)				
MEDLINE	20402566				
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L Division of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 141 row: A column: 11 Seq primer: 77 Class: BAC ends High quality sequence stop: 834.				
FEATURES	source Location/Qualifiers 1..834 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone="Plate=141 Col=11 Row=A" /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"				
BASE COUNT	270 a	132 c	162 g	270 t	
ORIGIN					
Query Match	72.8%; Score 18.2; DB 239; Length 834;				
Best Local Similarity	87.0%; Pred. No. 3.4e+02;				
Matches	20; Conservative	0;	Mismatches	3;	Indels 0; Gaps 0;
QY	2	caggtcgcgcgcagcatttcga	24		
Db	368	CAAGTCACAGCCAGCATTTCAA	346		
RESULT	11				
LOCUS	P858R	360 bp	DNA	GSS	25-JUL-2000
DEFINITION	Leishmania major Friedlin PAC P858 right end-sequence, similar to SW:VDA_BUCE 051940 L-ORNT1HNE 5-MONOOXYGENASE (PC. . . , N=107, Prob=1.9e-17, genomic survey sequence.				
ACCESSION	AL390537				
VERSION	AL390537.1	GI:9501513			
KEYWORDS	GSS.				
SOURCE	Leishmania major.				
ORGANISM	Leishmania major Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.				
REFERENCE	1 (bases 1 to 360)				
AUTHORS	Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and Smith,D.F.				
TITLE	A physical map of the Leishmania major Friedlin genome				
JOURNAL	Genome Res. 8 (2), 135-145 (1998)				

REFERENCE	TITLE	AUTHORS	JOURNAL	COMMENT
MEDLINE 98146435 2 (bases 1 to 360)	Direct Submission	Taylor R.G., Huckle, E.E.J., Ivens, A.C., Rajandream, M.A. and Barrell, B.G.	Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and allcats@sanger.ac.uk	see http://www.ebi.ac.uk/parsites/leish.html Details of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/projects/Lmajor/ The primer sequence can be obtained from allcats.sanger.ac.uk .
FEATURES	source	location/Qualifiers		
BASE COUNT	48 a	117 c	112 g	83 t
ORIGIN	1.360 /organism="Leishmania major" /strain="Friedlin" /db_xref="taxon:5664" /clone="PAC p858"			
Query Match	Best local Similarity	90.5%	Pred. No. 4.9e+02;	
Matches 19;	Conservative	0;	Mismatches 2;	Indels 0; Gaps 0;
QY	5	gtgcatcgcagcattcgat 25		
RESULT 12	AV186249/c	384 bp	mRNA	EST
LOCUS	AV186249	384 bp	mRNA	EST
DEFINITION	AV186249 yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo	Caenorhabditis elegans	CDNA clone yk497c6 5', mRNA sequence	
ACCESSION	AV186249			
VERSION	AV186249.1	GI:5568232		
KEYWORDS	EST.			
SOURCE	Caenorhabditis elegans.			
ORGANISM	Caenorhabditis elegans			
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitoidea			
AUTHORS	1 (bases 1 to 384)			
	Kohara, Y., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Mitsuki, H., Nishigaki, A., Motohashi, T., Zeng, Q., Watanabe, H., Sugimoto, A., San, M., Miyata, A., Mitani, Y., Iida, K., Uesugi, H., Sugiyama, Y. and Nomoto, H.			
TITLE	Expressed genes in C.elegans			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Yuji Kohara			
	Gene Library lab			
	National Institute of Genetics			
	Yata 1111, Mishima, Shizuoka 411, Japan			
	Tel: 81-559-81-6854			
	Fax: 81-559-81-6855			
	Email: ykohara@lab.nig.ac.jp.			
FEATURES	location/Qualifiers			
source	1.384			
	/organism="Caenorhabditis elegans"			
	/strain="N2"			
	/db_xref="taxon:6239"			
	/clone="yk497c6"			
	/clone_lib="yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo"			
	/sex="hermaphrodite"			
	/dev_stage="embryo"			
BASE COUNT	120 a	85 c	78 g	101 t
ORIGIN				

Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATTCGACTCCTATAGG
Class: BAC ends
High quality sequence stop: 437.
Location/Qualifiers

FEATURES

source

1..580
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nxb0039K21f"
/clone_lib="CUGI Rice BAC Library"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/note="Vector: pHELOBAC11; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."
BASE COUNT 194 a 111 c 105 g 168 t 2 others
ORIGIN

Query Match 71.2%; Score 17.8; DB 226; Length 580;
Best Local Similarity 90.5%; Pred. No. 5.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 4 ggtgcacgcacgacattcga 24
|||||
Db 112 GGTGCATCGCCACCTTTGCA 92

RESULT 16
BF203823 693 bp mRNA EST 06-NOV-2000
LOCUS 601968810F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:1111418 5',
DEFINITION mRNA sequence.
ACCESSION BF203823
VERSION BF203823.1 GI:11097409
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 693)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM99 row: c column: 03
High quality sequence start: 48
High quality sequence stop: 54.
Location/Qualifiers

FEATURES

source

1..693
/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="IMAGE:4111418"
/clone="NIH_MGC_17"
/clone_lib="NIH_MGC_17"
/tissue_type="rhodomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="organ: muscle; Vector: pCRB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Site selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 126 a 191 c 202 g 174 t
ORIGIN

Query Match 71.2%; Score 17.8; DB 145; Length 693;
Best Local Similarity 90.5%; Pred. No. 5.2e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 4 ggtgcacgcacgacattcga 24
|||||
Db 606 GGTGCATCCCGACCTTTGCA 626

RESULT 17
CNS003ZE 888 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence 17 end of BAC #
DEFINITION BACR09N17 of Rpet-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL065360
VERSION AL065360.1 GI:4942712
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 888)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named Rpet-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers

FEATURES

source

1..888
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT	ORIGIN
201 a	219 c
215 g	204 t
49 others	

Query Match	71.2%;	Score 17.8;	DB 219;	Length 888;
Best Local Similarity	90.5%;	Pred. No. 5.3e+02;		
Matches 19; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

RESULT	18
CNS058G5/c	
LOCUS	
DEFINITION	
<p>Tetradon nigriviridis genome survey sequence T3 end of clone 026108 of library A from Tetradon nigriviridis, genomic survey</p>	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

BASE COUNT	216 a	269 c	338 g	249 t	18 others
ORIGIN					

Query	Match	Similarity	Score	DB	Length
Best Local	71.2%	90.5%	17.8	222	1090
Matches	19	Conservative	0	Mismatches	2
				Indels	0
				Gaps	0
QY	5	gtagcatgcagcagcattccgat	25		
Db	989	gtgcattgcagcagcattccgat	969		

RESULT	19
AK018497/c	
LOCUS	
DEFINITION	
ACCESSION	AK018497
VERSION	1862 bp mRNA
KEYWORDS	Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone:9030409G11, full insert sequence.
SOURCE	AK018497.1 GI:12658226
	CAP trapper.
	Mus musculus. (strain:C57BL/6J) adult male colon cDNA to mRNA,

ORGANISM	REFERENCE
<i>Mus musculus</i>	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
1 (sites)	
Carninci, P., and Hayashizaki, Y.	
High-efficiency full-length cDNA cloning	
Methods Enzymol. 303, 19-44 (1999)	
2 (sites)	
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,	
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.	
Normalization and subtraction of cap-trapper-selected cDNAs to	
prepare full-length cDNA libraries for rapid discovery of new genes	
Genome Res. 10 (10), 1617-1630 (2000)	
JOURNAL MEDLINE	
20499374	
3 (sites)	
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,	
AUTHORS	

COMMENT

Rot = 185.2. Second strand cDNA was prepared with the primer
 adapter of sequence [5'
 GAGGAGAGATTCGACGATTATTAATTCACCCGCCCC 3']. cDNA was cleaved
 with BamHI and XhoI. Vector: a modified pBluescript KS(+) after
 bulk excision from Lambda FLX I. Cloning sites, 5' end: SalI; 3'
 end: BamHI. Host: DH10B.

FEATURES

source

1.1862
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /db_xref="MGI:1904075"
 /db_xref="MGI:1918779"
 /clone="9030409G11"
 /sex="male"
 /tissue_type="colon"
 /clone_11b="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="adult"
 BASE COUNT 390 a 540 c 564 g 367 t 1 others
 ORIGIN

Query Match 71.2%; Score 17.8; DB 192; Length 1862;
 Best local Similarity 90.5%; Pred. No. 5.7e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 caggtgcatgccagcatc 22
 ||||| 1 ||||| ||||| |||||
 Db 478 CAGGTGCTTCGCCAGCATTTTC 458

RESULT 20
 BE089771 317 bp mRNA EST 12-JUN-2000
 LOCUS
 DEFINITION QV0-BT0704-120500-224-g09 BT0704 Homo sapiens cDNA, mRNA sequence.
 BE089771
 VERSION BE089771.1 GI:8480206
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 AUTHORS 1 (bases 1 to 317)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R.,
 Nagai,M.A., da Silva W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 MEDLINE 20202663
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?l=st2-QV0-BT0704-120
 500-224-g09st3-2000-05-12&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 317.
 Location/Qualifiers

FEATURES
 source
 1..317
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_11b="BT0704"

/dev_stage="Adult"
 /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
 SmaI; A mini-library was made by cloning products derived
 from ORESSES PCR (U.S. Letters Patent application No. 196
 776 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."

BASE COUNT 99 a 77 c 78 g 63 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 163; Length 317;
 Best local Similarity 83.3%; Pred. No. 6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcaggtgcatgccagcatcga 24
 ||||| 1 ||||| 1 |||||
 Db 59 GCAGGTGATCACCAGCATTCGA 82

RESULT 21
 A0890634 418 bp DNA GSS 10-NOV-1999
 LOCUS
 DEFINITION HS-2204_A1_G03_T7C CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate-2204 Col=5 Row=M, DNA sequence.
 A0890634
 VERSION A0890634.1 GI:6346824
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 AUTHORS 1 (bases 1 to 418)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.hltsc.washington.edu
 Plate: 2204 row: M column: 5
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 418.

FEATURES
 source
 1..418
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_11b="Plate-2204 Col=5 Row=M"
 /sex="male"
 /note="Organ: sperm; Vector: pBel0BAC11; BAC Clones in
 E-Coli DH10B"

BASE COUNT 119 a 98 c 102 g 99 t
 ORIGIN

Query Match 70.4%; Score 17.6; DB 235; Length 418;
 Best local Similarity 83.3%; Pred. No. 6.1e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 caggtgcatgccagcatcgt 25
 || ||| ||||| ||||| ||||| ||

DB 144 CACTGCTTCGCCAGCATTTCCAT 121

RESULT 22
LOCUS A0516531/c
DEFINITION HS.5052_B1-D03-SP6E RPCT-11 Human Male BAC Library Homo sapiens genomic clone Plate-628 Col=5 Row=H, DNA sequence.
ACCESSION A0516531
VERSION A0516531.1 GI:4748679
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 450)
Mahaitas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
PROC. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
JOURNAL 99380589
MEDLINE
COMMENT Contact: Mahaitas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCT-11. For BAC library availability, please contact Pletier de Jong (pletier@ejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
Plate: 628 row: H column: 5
Seq primer: SP6
Class: BAC ends
FEATURES
Location/Qualifiers
Source 1..450
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate-628 Col=5 Row=H"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
BASE COUNT 107 a 105 c 94 g 142 t 2 others
ORIGIN

Query Match 70.4%; Score 17.6; DB 229; Length 450;
Best Local Similarity 83.3%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcaggtgcatgccagcattcga 24
||||| 1111111111111111
DB 251 GCAGCGGATCACCGAGATTTCGA 228

RESULT 23
LOCUS BF654469/c
DEFINITION BF654469 475 bp mRNA
ACCESSION BF654469
VERSION BF654469.1 GI:11919601
KEYWORDS EST.
SOURCE cow.
COMMENT 20-DEC-2000

ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 475)
Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid,W.W. and Keele,J.W.
TITLE Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@emall.marc.usda.gov
Single pass sequencing. Bases called and all-trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: CTTTTCCTCAGCAGCAG
Plate: 71 row: H column: 16
Seq primer: ATTAGTGACACTAG.
FEATURES
Location/Qualifiers
Source 1..475
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI; Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."
BASE COUNT 91 a 156 c 134 g 94 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 151; Length 475;
Best Local Similarity 83.3%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ccagtgatccagcagcattcga 25
||||| 1111111111111111
DB 474 CAGTGATCGCGCTTCATTAGAT 451

RESULT 24
LOCUS BG266805/c
DEFINITION BG266805 506 bp mRNA
ACCESSION BG266805
VERSION BG266805.1 GI:12970082
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 506)
Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221

Email: walbot@stanford.edu
Plate: 1000101 row: 6 column: 12.
Location/Qualifiers

FEATURES

1..506
/organism="Zea mays"
/db_xref="dbEST:605008C11.x1"
/db_xref="taxon:4577"
/clone_lib="1000 - Unigene I from Maize Genome Project"
/note="This library represents the unique ESTs found in the first round of EST sequencing at Stanford University for the maize genome project. Sequences are present from libraries 486, 487, 496, 603, 605, 606, 614, 618, 660, 683, 687, 707, and 945. Contigs were assembled using TIGR's CAP program and a representative EST from each contig was selected for the Unigene set. All singlets were also selected."

BASE COUNT 96 a 168 c 129 g 113 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 175; Length 506;
Best Local Similarity 83.3%; Pred. No. 6.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcagtgatcgccagcattcga 24
|||||
Db 456 GCAGTGATCGCCAGCATTCA 433

RESULT 25

AV619677 549 bp mRNA EST 15-DEC-2000
LOCUS AV619677 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
DEFINITION reinhardtii cDNA clone LC011f07_r 5', mRNA sequence.
ACCESSION AV619677
VERSION AV619677.1 GI:10768852
KEYWORDS EST
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 549)
Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y. and Tabata, S.
Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
DNA Res. 7 (5), 305-307 (2000)
20539644

JOURNAL MEDLINE
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1533-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers

FEATURES
source 1..549
/organism="Chlamydomonas reinhardtii"
/strain="c9"
/db_xref="taxon:3055"
/clone_lib="LC011f07_r"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimatized condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"

BASE COUNT 105 a 171 c 179 g 94 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 31; Length 549;
Best Local Similarity 83.3%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcagtgatcgccagcattcga 24
|||||
Db 338 GCAGGAGCAGCGCCAGCATTCAA 361

RESULT 26

AO795581 558 bp DNA GSS 04-AUG-1999
LOCUS AO795581
DEFINITION nbxb0056a12f CUBI Rice BAC Library Oryza sativa genomic clone
nbxb0056a12f, DNA sequence.
ACCESSION AO795581
VERSION AO795581.1 GI:5703966
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 558)
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
CONTACT: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCATCATAGG
Class: BAC ends
High quality sequence stop: 334.
Location/Qualifiers

FEATURES
source 1..558
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbxb0056a12f"
/clone_lib="CUBI Rice BAC Library"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
HindIII; Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 Kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9%. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."

BASE COUNT 122 a 177 c 139 g 120 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 233; Length 558;
Best Local Similarity 83.3%; Pred. No. 6.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 ccagtgatcgccagcattcga 25
|||||
Db 462 CCAGTGATCGCCAGCATTGAT 485

RESULT 27
 A1665214/c 582 bp mRNA EST 02-FEB-2000
 LOCUS 605008C11.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
 DEFINITION cDNA, mRNA sequence.
 ACCESSION A1665214
 VERSION A1665214.2 GI:4804415
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade: Panicoidae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 582)
 AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
 JOURNAL Unpublished (1999)
 COMMENT On May 11, 1999 this sequence version replaced gi:4776211.
 CONTACT: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 605008 row: C column: 11.
 FEATURES
 SOURCE
 1. 582
 /organism="Zea mays"
 /cultivar="Ohio43"
 /db_xref="taxon:4577"
 /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
 /tissue_type="nucellar, embryo, and endosperm"
 /dev_stage="10-14 days post-pollination"
 /lab_host="DH5(alpha)"
 /note="Organ: Kernel; Vector: PAD-GAL4-2'; Site: 1: EcoRI; Site-2: XhoI; Kernel endosperm cDNA library from Schmidt lab"
 BASE COUNT 115 a 191 c 146 g 130 t
 ORIGIN
 Query Match 70.4%; Score 17.6; DB 23; Length 582;
 Best Local Similarity 83.3%; Pred. No. 6.3e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 gcagtgcatgcgcagcattcga 24
 ||||||||| ||||||| |||
 Db 483 GCAGTGCAATCACGACGATACCA 460
 RESULT 28
 BF425225/c 615 bp mRNA EST 28-NOV-2000
 LOCUS Su5410.0.Y1 Gm-cl069 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 DEFINITION UB10U11TN-CONJUGATING ENZYME F2-17 KD 9 ;, mRNA sequence.
 ACCESSION BF425225
 VERSION BF425225.1 GI:11413214
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 615)
 REFERENCE 1 (bases 1 to 615)
 AUTHORS Shoemaker R., Keim P., Vodkin L., Erpelting J., Corryell V., Khanna A., Bolla B., Marra M., Hillier L., Kucaba T., Martin J., Beck C., Wille T., Underwood K., Steptoe M., Theising B., Allen M., Bowers Y., Person B., Swaller T., Gibbons M., Pape D., Harvey N., Schurk R., Rutter E., Kohn S., Shin T., Jackson Y., Cardenas M., McCann

TITLE
 JOURNAL
 COMMENT
 CONTACT: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
 High quality sequence stop: 412.
 FEATURES
 SOURCE
 1. 615
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl069-1123"
 /clone_lib="Gm-cl069"
 /tissue_type="degenerating cotyledons, 9-10 day old etiolated seedling"
 /lab_host="DH10B"
 /note="Vector: Bluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from mRNA isolated from degenerating cotyledons of 9-10 day old etiolated seedlings for the cultivar Williams. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."
 BASE COUNT 148 a 165 c 131 g 171 t
 ORIGIN
 Query Match 70.4%; Score 17.6; DB 148; Length 615;
 Best Local Similarity 83.3%; Pred. No. 6.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 2 caggtgcattgcgcagcattcga 25
 ||||||||| ||||||| |||
 Db 532 CAAGTGCAATCACGATTTGGAT 509
 RESULT 29
 CNS03VDY/c 1032 bp DNA GSS 18-MAY-2000
 LOCUS Tetradon nigroviridis genome survey sequence T7 end of clone
 DEFINITION 061N02 of library G from Tetradon nigroviridis, genomic survey sequence.
 ACCESSION AL262303
 VERSION AL262303.1 GI:7983929
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetradon.
 1 (bases 1 to 1032)
 REFERENCE 1 (bases 1 to 1032)
 AUTHORS Roest-Criolius H., Jallion O., Dasilva C., Pizames C., Fisher C., Bouneau L., Billault A., Quetier F., Saurin W., Bernot A. and Weissenbach J.
 TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
 JOURNAL Unpublished

REFERENCE	2 (bases 1 to 1032)									
AUTHORS	Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.									
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence									
JOURNAL	Unpublished									
REFERENCE	3 (bases 1 to 1032)									
AUTHORS	Genoscope.									
TITLE	Direct Submission									
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases									
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon .									
FEATURES	Location/Qualifiers									
SOURCE	1..1032									
	/organism="Tetraodon nigroviridis"									
	/db_xref="taxon:99883"									
	/clone="061N02"									
	/clone_11b="G"									
	/note="Genoscope sequence ID : COBG061D501P1-end : T7"									
BASE COUNT	214 a 296 c 324 g 194 t 4 others									
ORIGIN										
Query Match	70.4% ; Score 17.6; DB 221; Length 1032;									
Best Local Similarity	83.3% ; Pred. No. 6,7e+02;									
Matches	20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;									
Oy	2 caggtgcatgcgcgcatttcgat 25									
	11 1111111111111111									
Db	518 CATGTCATCGCCACCATCTCGAT 495									
RESULT	30									
CNS03B6D										
LOCUS	CNS03B6D 1039 bp DNA GSS 15-MAY-2000									
DEFINITION	Tetraodon nigroviridis genome survey sequence T7 end of clone 011022 of library G from Tetraodon nigroviridis, genomic survey sequence.									
ACCESSION	AL236110.1 GI:7895245									
VERSION	AL236110									
KEYWORDS	GSS: genome survey sequence.									
SOURCE	Tetraodon nigroviridis.									
ORGANISM	Tetraodon nigroviridis									
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Actinopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.									
REFERENCE	1 (bases 1 to 1039)									
AUTHORS	Roest-Crollius,H., Jalllon,O., Dasilva,C., Fizames,C., Fisher,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.									
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis									
JOURNAL	Unpublished									
REFERENCE	2 (bases 1 to 1039)									
AUTHORS	Roest-Crollius,H., Jalllon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.									
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence									
JOURNAL	Unpublished									
REFERENCE	3 (bases 1 to 1039)									
AUTHORS	Genoscope.									
TITLE	Direct Submission									
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases									
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .									
FEATURES	Location/Qualifiers									

```

source
1. 1039
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="011022"
/clone_11b="g"
/note="Genoscope sequence ID : COBG01BH11LP1-end : T7"

BASE COUNT      248 a      222 c      236 g      329 t      4 others
ORIGIN

Query Match      70.4%: Score 17.6; DB 220; Length 1039;
Best Local Similarity 83.3%: Pred. No. 6.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      1  gcaggtgcacgcacagcattcgca 24
        ||| || ||| ||| ||| ||| |||
Db      254  GCATGTTAATCGTCAGCATTTGCA 277

RESULT 31
BG027526/c      1077 bp      mRNA      EST      24-JAN-2001
LOCUS      602295377F1 NIH_MGC_86 Homo sapiens CDNA clone IMAGE:4390003 5',
DEFINITION      mRNA sequence.
ACCESSION      BG027526
VERSION      BG027526.1 GI:12416239
SOURCE      EST.
ORGANISM      human.
             Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS      National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE      1 (bases 1 to 1077)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cga@bbs-remail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LNLN at:
             http://image.llnl.gov
             Plate: LHAM10078 row: j column: 20
             High quality sequence stop: 620.
             Location/Qualifiers
               1. 1077
                 /organism="Homo sapiens"
                 /db_xref="taxon:9606"
                 /clone="IMAGE:4390003"
                 /clone_11b="NIH_MGC_86"
                 /tissue_type="osteosarcoma, cell line"
                 /lab_host="DH10B (phage-resistant)"
                 /note="Organ: bone; Vector: PCMV-SFOR6; Site_1: NotI;
                 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                 Average insert size 1.533 kb. Library enriched for
                 full-length clones and constructed by Life Technologies.
                 Note: this is a NIH_MGC library."

BASE COUNT      316 a      231 c      352 g      178 t
ORIGIN

Query Match      70.4%: Score 17.6; DB 172; Length 1077;
Best Local Similarity 83.3%: Pred. No. 6.7e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY      2  caggtgcacgcacagcattcgat 25
        ||| ||| ||| ||| ||| ||| |||
Db      446  CAGTGCATCTCCAGCATTTCCAT 423

RESULT 32

```

BF933544/c
 LOCUS BF933544 376 bp mRNA EST 22-JAN-2001
 DEFINITION CM1-NT0270-211200-661-903 NT0270 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF933544
 VERSION BF933544.1 GI:12350868
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 376)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Naldi, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?PL=CM1&L2=CM1-NT0270-211200-661-903&L3=2000-12-21&L4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 10
 High quality sequence stop: 376.
 Location/Qualifiers
 1..376
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NT0270"
 /dev_stage="Adult"
 /note="Organ: nervous_tumor; Vector: puc18; Site: 1: SmaI; Site: 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196/716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 77 a 104 c 99 g 96 t
 ORIGIN
 Query Match 69.6%; Score 17.4; DB 171; Length 376;
 Best Local Similarity 94.7%; Pred. No. 7.5e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 4 ggtgcatcgcagcattc 22
 ||||||||||||||||
 Db 329 ggtgcatcgcagcattc 311
 RESULT 33
 LOCUS AQ171736 527 bp DNA GSS 17-OCT-1998
 DEFINITION HS_3172_B2_G07_T7 CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=3172 Col=14 Row=N, DNA sequence.
 ACCESSION AQ171736
 VERSION AQ171736.1 GI:3569103
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 527)
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.
 Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
 TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 3172 Row: N Column: 14
 Class: BAC ends
 High quality sequence stop: 527.
 Location/Qualifiers
 1..527
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Plate=3172 Col=14 Row=N"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 148 a 125 c 135 g 117 t 2 others
 ORIGIN
 Query Match 69.6%; Score 17.4; DB 225; Length 527;
 Best Local Similarity 94.7%; Pred. No. 7.8e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 caggtgcatcgcagcatt 20
 ||||||||||||||||
 Db 185 caggtgcatcgcagcatt 203
 RESULT 34
 LOCUS AQ350006 634 bp DNA GSS 07-MAY-1999
 DEFINITION RPCI11-117D10.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-117D10,
 DNA sequence.
 ACCESSION AQ350006
 VERSION AQ350006.1 GI:4177341
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 634)
 Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.
 Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
 Unpublished (1997)
 Other_GSSs: RPCI11-117D10.TV
 Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbeetlgr.org
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@ejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html

Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers
1..634

ORGANISM
/organism="Homo sapiens"
/db_xref="GDB:7544625"
/db_xref="taxon:9606"
/clone="RPCT-11-117D10"
/clone_lib="RPCT-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPCT11 Human Male BAC library"
BASE COUNT
158 a 143 c 195 g 138 t
ORIGIN

Query Match
Best Local Similarity 94.7%; Score 17.4; DB 221; Length 634;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY
2 caggtgcacgcacgacatt 20
|||||
Db 179 CAGGTGCATGCCACGACATT 197

RESULT 35

CNS03HMM/c 786 bp DNA GSS 17-MAY-2000
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence T7 end of clone
027c16 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION
AL244481 GI:7965493
VERSION
GSS: genome survey sequence.
KEYWORDS
Tetraodon nigroviridis.
SOURCE
Tetraodon nigroviridis

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphae; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 786)

REFERENCE
Roest-Crollius, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C.,
Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished

TITLE
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL
REFERENCE
2 (bases 1 to 786)
Roest-Crollius, H., Jallion, O., Dasilva, C., Bonneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished

TITLE
3 (bases 1 to 786)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale genome sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

JOURNAL
COMMENT
Location/Qualifiers
1..786
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="027c16"
/clone_lib="G"
/note="Genoscope sequence ID : C08G027B808LPI-end : T7"

BASE COUNT
236 a 188 c 199 g 160 t 3 others
ORIGIN

Query Match
Best Local Similarity 94.7%; Score 17.4; DB 221; Length 786;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY
1 gcaggtgcacgcacgacatt 19
|||||
Db 495 GCAGGTGCATGCCACGACATT 477

RESULT 36

BE156980/c 147 bp mRNA EST 21-JUN-2000
LOCUS
DEFINITION
RC3-HT0371-100100-012-c01 HT0371 Homo sapiens cDNA, mRNA sequence.
ACCESSION
BE156980
VERSION
BE156980.1 GI:8619701
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 147)

REFERENCE
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R.,
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
Simpson, A. J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
JOURNAL
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC3-HT0371-100100-012-c01&t3=2000-01-10&t4=1>)
Seq primer: puc 18 forward
High quality sequence stop: 147.

FEATURES

source

Location/Qualifiers
1..147
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0371"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT
45 a 29 c 60 g 13 t
ORIGIN

Query Match
Best Local Similarity 86.4%; Score 17.2; DB 164; Length 147;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY
2 caggtgcacgcacgacattcg 23
|||||
Db 103 CAGGTGCATGCCACGACATTTCG 82

RESULT 37
LOCUS AA686065 309 bp mRNA EST 03-APR-1998
DEFINITION EST109111 Rat PC-12 cells, NGF-treated (9 days) Rattus sp. cDNA clone RPNAL21 5' end similar to Tyrosine hydroxylase, mRNA sequence.
ACCESSION AA686065
VERSION AA686065.1 GI:2672663
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 309)
Lee,N.H., Weinstein,K.G., Kirkness,E.F., Earle-Hughes,J.A., Fuldner ,R.A., Marmaras,S., Glodek,A., Socayne,J.D., Adams,M.D., Kerlavage ,A.R., Fraser,C.M. and Venter,J.C.
Comparative expressed-sequence-tag analysis of differential gene expression profiles in PC-12 cells before and after nerve growth factor treatment
Proc Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
JOURNAL MEDLINE
COMMENT Other_ESTs: TC45033
Contact: Lee, NH
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13 Reverse.
Location/Qualifiers
1..309
/organism="Rattus sp."
/db_xref="ATCC (inhost):2003572"
/db_xref="taxon:10118"
/clone="RPNAL21"
/note="Vector: Rat PC-12 cells, NGF-treated (9 days)"
/note="Vector: Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; poly(A)+ RNA was purified from 9-day NGF treated PC12 cells. cDNA was constructed using an oligo-dT primer and directionally cloned using the Lambda Zap II Vector kit by Stratagene"

BASE COUNT 66 a 85 c 84 g 67 t 7 others
ORIGIN

Query Match 68.8%; Score 17.2; DB 10; Length 309;
Best Local Similarity 82.6%; Pred. No. 9.2e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 caggtgcacgcacgacattcga 24
||||| ||||||| ||
Db 219 CAGGTGATCGCCAGCAGCTCGGA 241

RESULT 38
LOCUS A2273781/c 435 bp DNA GSS 26-JUL-2000
DEFINITION RPCI-23-107E20.TJ RPCI-23 Mus musculus genomic clone RPCI-23-107E20
ACCESSION A2273781
VERSION A2273781.1 GI:9487398
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 435)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akınret ,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
Mouse BAC End Sequences from library RPCI-23

JOURNAL
COMMENT Unpublished (1999)
Other_GSSs: RPCI-23-107E20.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (inforesgen.com). BAC end page: http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html
Plate: 107 row: E column: 20
Seq primer: Sp6
Class: BAC ends.
Location/Qualifiers
1..435
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-107E20"
/clone.lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site-1: EcoRI; Site-2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylation site selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 122 a 68 c 70 g 175 t
ORIGIN

Query Match 68.8%; Score 17.2; DB 240; Length 435;
Best Local Similarity 86.4%; Pred. No. 9.4e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcaggtgcacgcacgacattc 22
||||| ||||||| |||||||
Db 152 GCAGGTAATCGCTAGCATTC 131

RESULT 39
LOCUS BF291841 495 bp mRNA EST 17-NOV-2000
DEFINITION WHE2204_F02.L04Z5 Aegilops speltoides another cDNA library Aegilops speltoides cDNA clone WHE2204_F02.L04, mRNA sequence.
ACCESSION BF291841
VERSION BF291841.1 GI:11222905
KEYWORDS EST.
SOURCE Aegilops speltoides.
ORGANISM Aegilops speltoides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae ; Triticeae; Aegilops.
REFERENCE 1 (bases 1 to 495)
Akhunov,E., Anderson,O.D., Chao,S., Chin,A., Choi,D.W., Close,T.J., Fenton,R.D., Han,P.S., Hsia,C.C., Kang,Y., Kianian,P., Lazo,G.R., Miller,R., Otto,C., Rausch,C.J., Seaton,C.L., Simons,K., Tong,J.C. and Zhang,D.
The structure and function of the expressed portion of the wheat genomes - Another cDNA library from Aegilops speltoides
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105959773

Fax: 5105595818
Email: oanderson@pm.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Strategene SK primer.
Location/Qualifiers

FEATURES

source 1..495
/organism="Aegilops speltoides"
/cultivar="F2 from 2-12-4-8-1-1-1" x PI36909-12-811-(1)
/db_xref="taxon:4573"
/clone_lib="Aegilops speltoides anther cDNA library"
/tissue_type="anther"
/dev_stage="Premeiotic anthers"
/lab_host="E. coli SOLR"
/note="Vector: Lambda uni-ZAP XR, excised phagemid; Site 1: EcoRI; Site 2: XhoI; Plants were grown in a growth chamber at the University of California, Davis (Akhunov). Premeiotic anthers were harvested, total RNA and poly(A) RNA were prepared, from each tissue and then pooled. A cDNA library was made, and the cDNA clones were in vivo excised to give Bluescript phagemids in the T3 Close Lab (Akhunov, Chin, Choi, Close, Penton, Kianian, Otto, Simons, Zhang) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson Lab (all other authors)."

BASE COUNT

111 a 116 c 123 g 145 t

ORIGIN

Query Match 68.8%; Score 17.2; DB 146; Length 495;
Best Local Similarity 86.4%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 ggtgcatgcgcagcatcgcg 25
||||| ||||| ||||| |||||
Db 276 GGTGAGCGCCACATTCGAT 297

RESULT 40

TA47A090 519 bp DNA GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 47a09, reverse sequence,
DEFINITION genomic survey sequence.
ACCESSION AL455103
VERSION AL455103.1 GI:11856556
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE

1 (bases 1 to 519)

AUTHORS

Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.

TITLE

Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk

JOURNAL

COMMENT

Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T_brucei/.

FEATURES
source Location/Qualifiers
1..519
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="47a09"

BASE COUNT 129 a 128 c 129 g 133 t

ORIGIN

Query Match 68.8%; Score 17.2; DB 258; Length 519;
Best Local Similarity 86.4%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 ggtgcatgcgcagcatcgcg 25
||||| ||||| ||||| |||||
Db 490 GGTGATCGCGACGATTCGCT 511

RESULT 41

TA329B040/c 544 bp DNA GSS 13-DEC-2000
LOCUS T. brucei sheared genomic DNA clone 329B04, reverse sequence,
DEFINITION genomic survey sequence.
ACCESSION AL492386
VERSION AL492386.1 GI:11868648
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 544)
AUTHORS Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk

COMMENT

Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/projects/T_brucei/.

FEATURES

source

Location/Qualifiers
1..544
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="329B04"

BASE COUNT 143 a 125 c 158 g 118 t

ORIGIN

Query Match 68.8%; Score 17.2; DB 258; Length 544;
Best Local Similarity 86.4%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 4 ggtgcatgcgcagcatcgcg 25
||||| ||||| ||||| |||||
Db 217 GGTGATCGCGACGATTCGCT 196

RESULT 42

BF008069

LOCUS BF008069 555 bp mRNA EST 06-OCT-2000
 DEFINITION 1698729 Amblyomma americanum adult Lambda Zap Express Amblyomma
 americanum cDNA, mRNA sequence.
 ACCESSION BF008069
 VERSION BF008069.1 GI:10708344
 KEYWORDS EST.
 SOURCE Amblyomma americanum.
 ORGANISM Amblyomma americanum.
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 Parasitiformes; Ixodida; Ixodidae; Amblyomma.
 REFERENCE 1 (bases 1 to 555)
 Hill, C.A. and Gutierrez, J.A.
 Analysis of the expressed genome of the lone star tick, Amblyomma
 americanum (Acari: Ixodidae) using an expressed sequence tag
 approach
 Microb. Comp. Genomics 5 (2), 89-101 (2000) In press
 CONTACT: Hill CA
 Animal Science Discovery Research
 Elanco Animal Health, A Division of Eli Lilly and Company
 PO Box 708, 2001 West Main St., Greenfield, IN 46140, USA
 Tel: 317 277 0826
 Fax: 317 277 4522
 Email: HILL-CATHERINE.A@LILLY.COM.

FEATURES
 source
 1..555
 /organism="Amblyomma americanum"
 /db_xref="taxon:6943"
 /clone_lib="Amblyomma americanum adult Lambda Zap Express"
 /sex="Male, Female"
 /dev_stage="Adult"
 /note="Vector: pRK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 160 a 146 c 118 g 128 t 3 others
 ORIGIN

Query Match 68.8%; Score 17.2; DB 143; Length 555;
 Best Local Similarity 86.4%; Pred. No. 9.7e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcaagtcgacgcgcagcatttc 22
 ||||| ||||| ||||| |||||
 Db 249 GCGATACATGCTCGACATTTC 270

RESULT 43
 A212380 565 bp DNA GSS 09-JUN-2000
 LOCUS Sheared DNA-95G12.TF Sheared DNA Trypanosoma brucei genomic clone
 DEFINITION
 Sheared DNA-95G12, DNA sequence.
 ACCESSION A212380
 VERSION A212380.1 GI:8429281
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei.
 ORGANISM Trypanosoma brucei.
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma.
 REFERENCE 1 (bases 1 to 565)
 El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
 Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
 Fraser, C. and Adams, M.
 Determination of clone end sequences from Trypanosoma brucei GUTat
 10.1 sheared DNA library
 Unpublished (1999)
 Other GSSs: Sheared DNA-95G12.TR
 CONTACT: Najib M. El-Sayed
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0200
 Email: nelsayed@tigr.org
 Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
 DNA library constructed at TIGR. Clones will be available for

FEATURES
 source
 1..565
 /organism="Trypanosoma brucei"
 /strain="TRE0927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone_lib="Sheared DNA-95G12"
 /clone_lib="Sheared DNA"
 /note="Vector: pUC18; Site_1: SmaI; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from a cloned population of
 Trypanosoma brucei (TRE0927/4 GUTat 10.1) was mechanically
 sheared to give a tight size distribution (approx 2 kb).
 The v + i method used for the library construction is
 described in detail in Smith, H.O. and Venter, J.C.
 (Making small insert libraries for whole genome shotgun
 sequencing projects. In Genome Sequencing: A Practical
 Approach, eds. M. Vaudin and B. Barrell, Oxford University
 Press, 1999)."

BASE COUNT 170 a 123 c 136 g 136 t
 ORIGIN

Query Match 68.8%; Score 17.2; DB 239; Length 565;
 Best Local Similarity 86.4%; Pred. No. 9.7e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 ggtgcacgcgcagcattcgat 25
 ||||| ||||| ||||| |||||
 Db 496 GGTGATCGCGACGATTCGCT 475

RESULT 44
 AA769145 570 bp mRNA EST 08-FEB-1998
 LOCUS ob51n03.s1 NCI-CGAP_GCB1 Homo sapiens cDNA clone IMAGE:134933 3',
 DEFINITION
 mRNA sequence.
 ACCESSION AA769145
 VERSION AA769145.1 GI:2820383
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 570)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/dbp/image/image.html
 Insert length: 539 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 476.
 Location/Qualifiers
 1..570
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:134933"
 /clone_lib="NCI-CGAP_GCB1"
 /tissue_type="germinal center B cell"

Search completed: October 9, 2001, 13:46:43
 Job time: 9518 sec

/lab host="DH108"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marth (CBER). cDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-TGTTACCAATCTGAAGTGGAGCGCCGCTCATTTTCTTTTCTTTTCTTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 117 a 144 c 190 g 118 t 1 others
 ORIGIN

Query Match 68.8%; Score 17.2; DB 11; Length 570;
 Best Local Similarity 86.4%; Pred. No. 9.7e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 caggtgcatcgccagcatlcc 23
 ||||| ||||| |||||
 Db 212 CAGGTGCGACGCCAGCATTCG 191

RESULT 45
 BF008062 646 bp mRNA EST 06-OCT-2000
 LOCUS 1664118 Amblyomma americanum adult Lambda Zap Express Amblyomma
 DEFINITION americanum cDNA, mRNA sequence.
 ACCESSION BF008062
 VERSION BF008062.1 GI:10708337
 KEYWORDS EST.
 SOURCE Amblyomma americanum.
 ORGANISM Amblyomma americanum.

REFERENCE 1 (bases 1 to 646)
 AUTHORS Hill,C.A and Gutierrez,J.A.
 TITLE Analysis of the expressed genome of the lone star tick, Amblyomma
 americanum (Acari:Ixodidae) using an expressed sequence tag
 approach

JOURNAL Microb. Comp. Genomics 5 (2), 89-101 (2000) In press
 COMMENT Contact: Hill CA
 Animal Science Discovery Research
 Elanco Animal Health, A Division of Eli Lilly and Company
 PO Box 708, 2001 West Main St., Greenfield, IN 46140, USA
 Tel: 317 277 0826
 Fax: 317 277 4522

FEATURES
 source
 1..646
 Location/Qualifiers
 /organism="Amblyomma americanum"
 /db_xref="taxon:6943"
 /clone_lib="Amblyomma americanum adult Lambda Zap Express"
 /sex="Male, Female"
 /dev_stage="Adult"
 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 154 a 153 c 172 g 164 t 3 others
 ORIGIN

Query Match 68.8%; Score 17.2; DB 143; Length 646;
 Best Local Similarity 86.4%; Pred. No. 9.8e+02;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gcagtgcatcgccagcatlcc 22
 ||||| ||||| |||||
 Db 554 GCAGATACATCGTCAGCATTTTC 533

Wed Oct 10 07:44:46 2001

us-09-396-196f-6.std.rst

Page 22

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 12:12:46 ; Search time 1666.31 seconds
(without alignments)
232.066 Million cell updates/sec

Title: US-09-396-196f-7

Perfect score: 25
Sequence: 1 gcatcgccagcatctgcatctcgt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 773874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_hgo_hum:*
20: em_hgo_iny:*
21: em_hgo_rod:*
22: em_hgo_hum1:*
23: em_hgo_hum2:*
24: em_hgo_hum3:*
25: em_hgo_hum4:*
26: em_hgo_hum5:*
27: em_hgo_hum6:*
28: em_hgo_hum7:*
29: em_hgo_hum8:*
30: em_hgo_iny1:*
31: em_hgo_iny2:*
32: em_hgo_other:*
33: em_hgo_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_hgt1:*
61: gb_hgt2:*
62: gb_hgt3:*
63: gb_hgt4:*
64: gb_hgt5:*
65: gb_hgt6:*
66: gb_hgt7:*
67: gb_hgt8:*
68: gb_hgt9:*
69: gb_hgt10:*
70: gb_hgt11:*
71: gb_hgt12:*
72: gb_hgt13:*
73: gb_hgt14:*
74: gb_hgt15:*
75: gb_hgt16:*
76: gb_hgt17:*
77: gb_hgt18:*
78: gb_hgt19:*
79: gb_hgt20:*
80: gb_hgt21:*
81: gb_hgt22:*
82: gb_hgt23:*
83: gb_hgt24:*
84: gb_hgt25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_pr10:*
95: gb_pr11:*
96: gb_pr12:*
97: gb_pr13:*
98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	DB ID	Description
1	25	100.0	1041 9	AR029499 Sequence
2	25	100.0	1041 9	AR034916 Sequence
3	25	100.0	1084 9	A11530 B10B gene o
4	25	100.0	1121 10	E00893 Genomic DNA
5	25	100.0	5793 2	J04423 E.coli 7, 8-
6	25	100.0	5872 9	A38246 Sequence 1
7	25	100.0	5872 9	A38251 Sequence 6
8	25	100.0	5872 9	A93674 Sequence 1

```

9      25 100.0 5872 9 A93679 Sequence 6
10     25 100.0 5872 9 AR101809 Sequence
11     25 100.0 5872 9 AR101810 Sequence
12     25 100.0 11022 1 AE000180
13     25 100.0 13801 1 AE005258
14     25 100.0 297816 2 AE002553
15     24 96.0 5526 2 AE250776
16     20.4 81.6 10593 1 AE003860
17     19.2 76.8 965 2 AE250770
18     19.2 76.8 4000 2 AE250768
19     19.2 76.8 62522 77 AC090154
20     19.2 76.8 62522 77 AC090154
21     19.2 76.8 68419 64 AC016014
22     19.2 76.8 116334 76 AC084172
23     19.2 76.8 186616 70 AC026294
24     18.6 74.4 80954 62 AC011276
25     18.6 74.4 112808 73 AC068024
26     18.6 74.4 137599 72 AC055855
27     18.2 72.8 1041 3 EMB107P2
28     18.2 72.8 3291 14 D88424
29     18.2 72.8 3451 10 E13998
30     18.2 72.8 4304 3 KFU31464
31     18.2 72.8 10239 1 AE004457
32     18.2 72.8 11382 3 RLE310723
33     18.2 72.8 11840 1 AE004541
34     18 72.0 128 2 EMB107P2
35     17.8 71.2 3748 3 MSU31280
36     17.8 71.2 3748 56 CVU39574
37     17.8 71.2 12107 1 AE002396
38     17.8 71.2 14645 65 AC018306
39     17.8 71.2 106462 83 AP003375
40     17.8 71.2 110000 84 TBCHRNA_03
41     17.8 71.2 183439 4 AC007809
42     17.8 71.2 195767 3 NMA722491
43     17.8 71.2 232744 5 AE003705
44     17.8 71.2 237119 65 AC017740
45     17.8 71.2 257867 4 AC005557

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D,Andrew
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
source Location/Qualifiers
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match 100.0%; Score 25; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcatgcgcagcattcgatccgct 25
Db 90 GCATGCCAGCATTTGCATCTCGT 114
RESULT 2

```

```

AR034916
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D,A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
source Location/Qualifiers
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcatgcgcagcattcgatccgct 25
Db 90 GCATGCCAGCATTTGCATCTCGT 114

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION B10B gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Bacterla; Proteobacteria; gamma subdivision; Enterobacteriaceae;
JOURNAL Escherichia.
FEATURES
source Location/Qualifiers
1.1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24.1064
/gene="B10B"
24.1064
/gene="B10B"
24.1064
/gene="B10B"
/codon_start=1
/transl_table=11
/protein_id="CAA00965.1"
/db_xref="GI:490219"
/db_xref="SWISS-PROT:P12996"
/translation="MAHRPMTLSQVTELEFKPLDILFEAOQVHRHDFPROVST
ILSTKACPECKCYKQSSRYKTELEERMEYQVLESRRKAKAGSTFCGCAAM
KNPHERDMPLYEOMYQYKAMGLEACMTLTGLSSQAKRLNNAAGLDIYNHLDTSPE
IGNITITRYSQRLDTEKRVDAKIVCSGSIYIGETVKDRAELIQLALPTPPES
VPIINLAVKGIPLADNDVDADFIRITIAVARIMPTSVRLSAGREONNEQ7QAMC
FMAGANSITFYGCKLITTPNPREDKDLQIFRKILGNPOTAVIADNEQORLEQALMT
PDTDEYVNAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcatgcgcagcattcgatccgct 25

```


Db 113 GCATGCCGACATTGATCTCGT 137

RESULT 4
LOCUS E00893 1121 bp DNA PAT 29-SEP-1997
DEFINITION Genomic DNA encoding biotin Synthetase.
ACCESSION E00893
VERSION E00893.1 GI:2169154
KEYWORDS JP 1986149091-A/1.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

REFERENCE 1 (bases 1 to 1121)
AUTHORS Hirono,Y., Kojima,T. and Kimura,H.
TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN
JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
NIPPON SODA CO LTD

COMMENT OS Escherichia coli
PN JP 1986149091-A/1
PD 07-JUL-1986
PF 24-DEC-1984 JP 1984272605
PI HIRONO YOSHIIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N1/5:00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
C12R1:19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=Escherichia coli Ns101;
CC Feature is identified by experimental;
FH Key Location/Qualifiers
FT CDS 42..1079
FT /product=biotin synthetase'.

FEATURES
source Location/Qualifiers
1..1121
/organism="Escherichia coli"
/db_xref="taxon:562"

BASE COUNT 289 a 296 c 325 g 211 t
ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 1121;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacggcagcatttcgactcgt 25
|||||
Db 131 GCATGCCGACATTGATCTCGT 155

RESULT 5
LOCUS E00893 5793 bp DNA BCT 28-FEB-1994
DEFINITION E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc protein, and dehydrobiotin synthetase (bioD), complete cds.
ACCESSION J04423
VERSION J04423.1 GI:145422
KEYWORDS 7,8-diamino-pelargonic acid aminotransferase;
7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
bioc gene; bioD gene; bioF gene; biotin synthetase; dehydrobiotin synthetase.

SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

REFERENCE 1 (bases 1 to 5793)
AUTHORS Otsuka,A.J., Buoncrisiani,M.R., Howard,P.K., Flamm,J. and Johnson,O.

TITLE The Escherichia coli biotin biosynthetic enzyme sequences
JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
MEDLINE 89066784
COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.
FEATURES
source Location/Qualifiers
1..5793
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
/comment=(98..574)
/note="ORF 1"
/codon_start=1
/transl_table=1
/protein_id="AA23513.1"
/db_xref="GI:455168"
/translation="MKLISNDLNDGDKLPRHYVFNMGNGYDGNISPHLADDPACTK
SFVYTCYDPAPTGSGMMHWYVNLPAIDTVLPLQGGSGLVAMPDGLVLRIDFGKGTG
YDGAAPPKGETHRYIFVHALDIERIDVDEGASAMVGFVHFHSLASITAMFS"
complement(633..1925)
/gene="bioA"
complement(633..1925)
/codon_start=1
/transl_table=1
/protein_id="AA23514.1"
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AA23514.1"
/db_xref="GI:457106"
/translation="MTTDDLAFDGRHIMPTYSNTSPLPYVYVSAEGCELIIISDGR
LVDGSSMMAAINGYVNHNPOLNAAMKSOIDAMSHVMEGIIHAPAIELCRILVMSGN
ALFECVPLADSGSVENVAEMKMAIOLYOMAKGEARQRELPFNNGYDGFEGAMSCDGN
SMHSLMGVYLPENLFAPAPOSRMDGEMDMGCFARLMAHREIAVLIIEPIVQCA
GGMHNPBWKRTKRICDREGILLINDEITATGGRGPKIPACHEAFLADIIICAKA
LITGGTMTLSATLITREVAETISKEACGFMHGPTPKNPLACAAANSIAIIESGMO
QOVADIEVQLEQOLAPARADEMADVAVLGAIGVETTHPVNMAALOKFEVEGVMIR
PFGRLIYLMPEYIILPOOLRLTAANRAVODETFQO"
2012..3052
/gene="bioB"
2012..3052
/codon_start=1
/transl_table=1
/product="biotin synthetase"
/protein_id="AAA23515.1"
/db_xref="GI:145425"
/translation="MAHRPRTLSQVTELEPKPLDLLFEAOYVHRQHPDROVST
LLSIKTGACPEDCRYCQPTSYKTLGLEAERLMEVEQVLESARKAKAGSTRFCGAM
KNPHERMDPYLEOMVOGVKAGLEACMTLGLTSSQORLANAGLDYVNNHLDTPSP
YGNITTRTYOERLDITLEKVRDAGIKVCSGIVGIGETVXRAGILQLQNLPTPPS
VPINMLVVKGTPLADNDVDAPFEIRIIVAVRIIMPITSYVRLSAGREOMNEQOAMC
EMAGANSIFVCKILLTTPREDKDLQFRILGILNPQOTAVLADNDNQOORLEQALMT
PDDETVAAL"

gene

CDS

/transl_table=1
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AAA23516.1"
/db_xref="GI:145426"

/translation="MSMOEKINAAIDARRADALRRRYVAGAGRWLVADROYLNF
SSNDYLSHHPOLIRAMOGAEQFGLSSGSSGVSVYHOLEEELAEMLGLSRA
LLFISGRANQAVIAAMAKRDRITADSLTASISILBAASISPSOLRRFANDVHRLR
LIASPCPOQNAVTEGVSMGDSAPLAETQVYQOHNGMLVNDAGTCVYIGGCGG
SCWILQKVRPELLVTEGKGFVSGAIVCSSTVADYDLQFARHLITSTSPRAQAL
BSLAVIIRSDSDGDAERKLAALITFRFAGVODLPFTLADSCSAIQPLIVGNSRALOL
AEKLRQCGCWYAIIRPVPAGTARLRLTLTAHMODIDRLLEVLHGNG"

4190..4945
/gene="bioc"
4190..4945
/gene="bioc"
/gene="bioc"

gene

CDS

/gene="bioc"

```

/codon_start=1
/transl_table=1
/protein_id="AA23517.1"
/db_xref="GI:145427"
/translation="MATVNNKQAIIAAFGRAAHYEQHADLROSADLLAMLQPKY"
HYLDAGCGPGMSRHHRRKQVTAALDISPMLVOAKQKADADHYLAGDIESLPLATA
TFDLAMSNLAVQWGNLSTALRELYRVRKGVAFITLVQGSLEPERHQAQVADERP
HANRFLPDELEOSLNGVYOHILQPTLWFDLMSRSLKIGATHLEGRDRL
TRSQLRLQALAMPQOGRPLTYHLFLGVIARE"
4938..5597
/gene="biob"
/locus_tag="BIOB"
/feature_key="biob"
4938..5597
/gene="biob"
/codon_start=1
/transl_table=1
/product="dethiobiotin synthetase"
/protein_id="AA23518.1"
/db_xref="GI:145428"
/translation="MSKRYVTGTDITVGTAVSCALLQAKRAGRTAGYKPVASG
EKTEGLRNSDALQKNSSLQIDYATVNPYTAEPSPHISAQEGRPESLVMGAG
LRALEQQAQVWLVGAGWETPLSDITFADWVQEQLPVILVGVKLCINHAMLTA
QVIOHAGITLAGVANDVTPPKRHAEMYTLTLMIPRCWERSPGIQRKROPESS
T"
BASE COUNT 1363 a 1554 c 1631 g 1245 t
ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcatcgacgacattcgatctctgt 25
Db 2101 GCATCGCACGACATTTCGATCTCTGT 2125

RESULT 6
LOCUS A38246 5872 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION A38246
VERSION A38246.1 GI:2294844
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 1 14-Apr-1994;
LONZA AG (CH)
COMMENT
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 85016947 960227.
LOCATION/Qualifiers
FEATURES
Source
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
-35_signal 23..28
-10_signal 45..49
/evidence_name="PROMOTER PTAC"
/evidence=experimental

RBS 105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
/locus_tag="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/number=1
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAA02325.1"
/db_xref="GI:2294845"
/translation="MAHRPRWTLISQVTELEFEKPLDILFEAQVYHROHFPDROVYST
LISIKTGACPECKCYCPOSSRYKTCLEAEKRLMEVEYVLESARKAKAAGSTRPCGAAM
KPHERDMPLEOMVOGAKAMGLECKMTGLTLESQAOURLANAGLDYTNHMLDPSPER
YGNILITRTYQERLDTLEKVDAGIKVSGGIVGERTVXRAGILLQALMLPPES
VPINMLVKKGTPLADNDVDADFIRITAVARIMPTSYVRLSAGREOMEQTOAMC
PMAAGNSIYCGCKLTTPPEEDKDLOLFRLKGLNPOQTAVIAGDNEQORLEQALMT
PPTDEVYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATVNNKQAIIAAFGRAAHYEQHADLROSADLLAMLQPKY
HYLDAGCGPGMSRHHRRKQVTAALDISPMLVOAKQKADADHYLAGDIESLPLATA
TFDLAMSNLAVQWGNLSTALRELYRVRKGVAFITLVQGSLEPERHQAQVADERP
HANRFLPDELEOSLNGVYOHILQPTLWFDLMSRSLKIGATHLEGRDRL
TRSQLRLQALAMPQOGRPLTYHLFLGVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTTDLAFQDRHIMPTYSMTSPVYVVSAGCELIISDGR
LYDQSSWMAIHGYNHNPQIQAANKSODAMSHVPGCTTHAPATELCRLVATTPD
LECFYLDGSSVAEVPAKMAIQLQWAKGEARQRLTFKNGYHEDTGAHSVCDPDS
HMSLWKGYLPENLPAFAPOSRMDGEWBERDVGARLMAARHREIAVILEPYOGAG
GMRMYHPEMLKIRIKICDREGIILIADEIATGEGRTKFLPACETAETAPDILGKAL
TGGMTLSATLTTRVAVETISNGEAGCFMGPTPMGNPLCAANASLAILLESQMO
QVADIEVOLREOLAPARDAEMVADVAVRIGAIQVETHPVNNMALQKFEYEGWIRP
FGKLIYMPYIILIPQOLRLTAAVNNAVODETFQO"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/number=6
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6

```

```

/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDLRDGLKPHRHVFGMGYDGNISPHLAMDVPAGTK
SFVVCYDDAPDPTGSMHMYVNLPAIDRVLPQGFSGLYAMPDGVQTRIDFGKTG
YDGAAPKGEFTHRYFTYHALDIERLIDDEGASGAMVGFNFHFLASASTAMRS"
stem_loop
5583..5605
5583..5644
terminator
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best local Similarity 100.0%; Pred. No. 0.29;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 gcatgccagcattcgatccctg 25
|||||
Db      206 GCATGCCAGCATTTCGATCCTCGT 230

RESULT  7
A38251      5872 bp      DNA      PAT      05-MAR-1997
LOCUS      Sequence 6 from Patent WO9408023.
A38251
VERSION    A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL     Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)
COMMENT     Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42085 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 85016947 960227.
Location/Qualifiers
FEATURES
source      1..5872
location/Qualifiers
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=1
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWQEKINALLARRADALRRRYPYAGAGRWLVADDDOYINF
SSNDYGLSHHPQILIRAMQGADEQFGISGSGSVSYVHQALEELAELEWGLYSFA
LIIFSGFANQAVTAAAMAKEDRIIAURLSHASLLEAASLSPSOLRFRFANDVTHLAR
LIASPCGGQOMVYEGVFSMDGSAPLAEIQVQOHNQWLMVDDAHGTGCVIGEOGNG

```

```

SCWLKVPPELLVTFGKGFVSGAAVLCSSTVADYLLQFARLLYSTSPPAQAL
RASLAVIRSDGDARREKLALITRFAGVODLPFTLADSCSAIQPLIVGNSRALOL
AEKLRQGCWWTARPPPTVAGTARLLTLTAAHEMODDRLLEVLHGNG"
RBS      3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
CDS
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=1
/number=4
/evidence=experimental
/product="PTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYEYGTDFEYGTVASCALLQAAKAGRYTAGYKPYVSGS
EKTEPGLNSDALALGRNSSSLQDYATVNPYTFAPSPHIIISAQGRPIESLYMSG
LRALFQADWYLVGACGMFPLSDFTFADWMTQBOPLVLYVGRKIGGINAMLTA
QVIOHAGITLIGWYANDVTPPGKRAHEMTTLTMTAPPLIGETPMIAENPENATGK
YINLAFVDSYTLGFTSRL"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best local Similarity 100.0%; Pred. No. 0.29;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 gcatgccagcattcgatccctg 25
|||||
Db      206 GCATGCCAGCATTTCGATCCTCGT 230

RESULT  8
A93674      5872 bp      DNA      PAT      22-JAN-2000
LOCUS      Sequence 1 from Patent EP0798384.
A93674
VERSION    A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE   1 (bases 1 to 5872)
AUTHORS    Birch,O. and Brass,J.
TITLE      Biotechnological method of producing biotin
JOURNAL     Patent: EP 0798384-A 1 01-OCT-1997;
LONZA AG (CH)
Location/Qualifiers
FEATURES
source      1..5872
location/Qualifiers
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1..96
/promoter
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..48
-10_signal
45..49
RBS      105..119
/standard_name="BIOB RBS NO.9"
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
CDS
/codon_start=1

```

```

/transl_table=11
/evidence-experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPRTLSQVTELEFRKPLDLLFEAQOVHRQHPDFROVQVST
LISTGACPECKTCPOSRYKTGLEERLMEVEQVLESARKAKAGSTRFCMGAM
KNPERDMPYLEQVQVQKAMGLEAKLTGLTSSQOARLANAGLDVYNNILDTSPF
YGNITTRTYOERDITLEKVRADGIVKCSGIVGLGETVDRKADILLQANLPTPES
VPINMLYKGTPLANDNDVDADPEIRI IAVARIAMPTSYRLSAGRQNMEOQAMC
FMAGANSIFYGCKLITTPNPEEDKDLQFRKLGINPQOTAVLAGDNQOORLEQALM"
PDTDEYVNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELONYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVNCQALIAARFRAAHYEQHADLQROSAQDLMLLQRYKT
HYDAGCGPGMSRHRERKAOYTAIDLSPMLVOAKQADHYTLAGDIESLPLATA
TFDLAWSIAVQKCNLTALRELRYRVRGVAFVTLVQSLPELHQMQVDEHP
HANFELPPEDEISLNGVYOHHTQPTTLMFEDALSMRSLKIGATHLHEGRDPRIL
TRSQLQRLQALAMPQOQGRYPLTYHLFLGVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANONATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAPDOHRIWHPTSMTPSPYVYVSAAGCELISDGR
LYDQSSWMAIHGYNPOLNAAMKSIDAMSHVMCGITHAPTELCRLVAMTPOF
LECYTLADSGVAEYAMKALOYWOAKCARQRFITFRNGTIGDTFGASVCDPDS
MHSIMKCYLPENLPAPPOSMDGEWDERMVGPARLMAHREHIAVILIEPTVQAG
GMRYRHEWLRKRIKICDREGILLIADDELATGFRGKLFACHEAIADPILICGAL
TGGTMTLSATLTREVAETISNGEAGCFMHGPTFMGNPLACAANASILLESQDQO
OVADIEVOLREQLAPADAEMVADVRLCAIGVETTHPYNMAALQKPFVEQGVWIRP
FGKLIYMPYIILPQOLQRLTAANVRADVDFEFCO"
5088..5100
/standard_name="ORF1 RBS"
5098..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGDKLPHRHVFNMGVYDGNISPHLAMDVPAGTK
SEVVCYDPAPTGSGMWHVYVNLPAQDRVLPQCGSLVAMPDGVLOTDFDCKTG
YDGAAPKGETHXYITVIALDERIDVDEGASGAWGFWHSHLASASTIAMS"
5583..5605
stem_loop
terminator
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1552 C 1695 G 1307 T

```

```

ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 gcatcgccagcattcgatcgct 25
Db 206 GCATCGCCAGCATTTCCATCTCCT 230

RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch,O. and Brass,J.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
source
location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQETINAAIDARRADALRRYPVAVOGAGRMVLVDDROYLNF
SSNDYLGSHHPOLIRAMQCGARPOGCGSGGSHVSGSYVHQALEBELAEMLCYSRA
LLFTSGPAAHQVIAAMAKDRIRADRLSHASILLEASISPSQLRRFAHNDVTHLR
LIASPCQDQWVTEGVFSMDGSAPIAEIQOVQOHNQWLMVDAGTGVLEGQNG
CSWLQKVAPELLVYTFEGKFGVSGAAVYCSSTVADYLLQFARHLIYSTSMPPAQAL
BSLAVIRSDGDAAREKIALILTRFRAGVODLPTTLADSCSAIOPITVDNSRALDL
AEKLROQGCWVTAIRPPTVPAGTARLRLLTIAHEMDIDRLLEVLHGNG"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="DTB SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYEYTGDTVEGTVASCALLQAAKAGYRTAGYKPYVSGS
EKTPGELRNSDALALQNSSTQLDYATVNVYTAEPSPHITISAOGRPLIESVMSG
LRAEQADWVIVGAGCGWTPRLSDTTFPDDWVTOEDLPILVYGVKLGCTINAMTLA
OVIOHAGLTLAGVANDVTPPGKRHAEYMTTLTMTIPAPILGETPWLAEPEANAATGK

```

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 YINIAFVASTLGFTSRL"
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.29;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatgccagcattcgatcctgt 25
 Db 206 GCATGCCAGCATTTCGATCCTCGT 230

RESULT 10
 ARI01809
 LOCUS ARI01809 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 1 from patent US 6083712.
 ACCESSION ARI01809
 VERSION ARI01809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 FEATURES Location/Qualifiers

source 1..5872
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.29;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatgccagcattcgatcctgt 25
 Db 206 GCATGCCAGCATTTCGATCCTCGT 230

RESULT 11
 ARI01810
 LOCUS ARI01810 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION ARI01810
 VERSION ARI01810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
 FEATURES Location/Qualifiers

source 1..5872
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.29;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatgccagcattcgatcctgt 25
 Db 206 GCATGCCAGCATTTCGATCCTCGT 230

RESULT 12
 AE000180
 LOCUS AE000180 11022 bp DNA BCT 01-DEC-2000
 DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.

ACCESSION AE000180 000096
 VERSION AE000180.1 GI:1786988
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 9742617
 PUBMED 9278503

REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner,F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner,F.R.
 TITLE Direct Submission

REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett,G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 5 (bases 1 to 11022)
 AUTHORS Plunkett,G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

COMMENT
 This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@marlab.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). ***The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase "b" assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES
 source
 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"

```

repeat_region /db_xref="taxon:83333"
               <1..130
               /note="REP (repetitive extragenic palindromic) element;
               contains 4 REP sequences"
gene           complement(147..1430)
               /gene="ybhc"
               /note="b0772"
               complement(147..1430)
CDS            /gene="ybhc"
               /function="orf; Not classified"
               /note="1427; 98 pct identical to fragment YBHC_ECOLI
               SW:P46130 (300 aa) but contains 127 additional C-terminal
               residues"
               /codon_start=1
               /transl_table=11
               /product="putative pectinesterase"
               /protein_id="AAC73859.1"
               /db_xref="GI:1786989"
               /translation="MNTFVSRLALALAFVTLTACSTPPDORSPDOTAPGTSRPI
               LSAEAOHFDAOHFASLTGCAAMNPSPTLPAPDPFVPGPACTGCTHTTTIOAAD
               AAIKRTKROYIIVMGEYQGYVYPAAPGGITLYTGKPKIDVKIGLSIDGMSPA
               DWRHDVNPGRKTPGKPAWYDSCQSRSDISGLCAVFSQNGIOLQNLIENT
               LGSDVDAGNHPAVALRTDGDVQINNVIILGRONTPEFVNSQNLRTKROPTLV
               NSYIEGDVIDVSGCAVVFDFEYRVNSRTQCAVYEAAPATLSNIYGFVAVNSRN
               AFGGVQOLGRSLVDANTNGOVYIRDSATNEGNTAKPMADAVISMPPAGNTGSVD
               DNDTORLNDNTNRMMEVNNRGVSKVAEAKK"
               complement(1507..1535)
promoter       /note="factor Sigma70; predicted +1 start at 806574"
               /complement(1582..2058)
gene           /gene="ybhb"
               /note="b0773"
               complement(1582..2058)
CDS            /gene="ybhb"
               /note="b0773"
               complement(1582..2058)
               /function="orf; Unknown"
               /note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
               /codon_start=1
               /transl_table=11
               /product="orf, hypothetical protein"
               /protein_id="AAC73860.1"
               /db_xref="GI:1786990"
               /translation="MKLISNDLRDGLKPLRHRYHNGNGYGDNISHLAMDVDPAGTK
               SYVTGYDPDAPRGSSGMHWYVNLPRDTRVLPGGSGGLVAMPDGLQTRIDFGTG
               YDGAAPKGETHRIETVHALDIERDIVDGSAGAVGFVHPSHLASITMFS"
               complement(2108..2124)
protein_bind   /note="central position to predicted promoter:85..5"
               /bound_moiety="Rnas predicted site"
               complement(2117..3406)
gene           /gene="bioA"
               /note="b0774"
               complement(2117..3406)
CDS            /gene="bioA"
               /note="b0774"
               complement(2117..3406)
               /EC_number="2.6.1.62"
               /function="enzyme; Biosynthesis of cofactors, carriers:
               Biotin"
               /note="f429; 100 pct identical to BIOA_ECOLI SW: P12995"
               /codon_start=1
               /transl_table=11
               /product="7,8-diaminopelargonic acid synthetase"
               /protein_id="AAC73861.1"
               /db_xref="GI:1786991"
               /translation="MTTDLAFDORHIMHPYMSPTLPYVYVSAEGEELISDGR
               LVNGSSWAATIGYVNHROPOLNAAKSDITAMSHVFMFGITTHAPALELCKRLVAMPPO
               LECVFLADSSGSAVEYAMKALQYQAKGEARLTFPNGHGFPGMSVCDPDS
               MSLMKGYLENFAPAPQSRMDGWDMDVGFARLMAHHEIAVYIEPIYOGAG
               GMRMYHPMLKRIKICDREGILLADELATGFGTKLFACEHAETIADIDICLAKL
               TGGTMTLSATLTTRTEVAETISNGEACFGPMGKPLACAAANSLATLESQMOO
               OVADIEVOLREOLAPARDAEMVADVIRGATGVETTHVYVNAALOKFFVEGVNIRP
               FGLIYIMPYIITPOOLITRAAVRAVODETFPGQ"
               complement(2193..2221)
promoter       /gene="bioA"
               /note="factor Sigma70; predicted +1 start at 807260"
               /protein_id="AAC73862.1"
               /note="central position to bioB promoter: -20"

```

```

protein_bind   /bound_moiety="BioB documented site"
               3411..3450
               /note="central position to predicted promoter: -20; genetic
               evidence for the site"
protein_bind   /bound_moiety="BioB predicted site"
               complement(3411..3450)
               /note="central position to predicted promoter:50"
               /bound_moiety="BioB predicted site"
               complement(3411..3450)
               /note="central position to bioA promoter:50"
               /bound_moiety="BioB documented site"
               3413..3441
               /note="factor Sigma70; promoter bioB; documented +1 at
               808525"
               complement(3447..3473)
               /note="factor Sigma70; promoter bioA; documented +1 at
               808515"
gene           /gene="bioB"
               /note="b0775"
               3493..4533
               /note="b0775"
               3493..4533
               /gene="bioB"
               /EC_number="2.8.1.-"
               /function="enzyme; Biosynthesis of cofactors, carriers:
               Biotin"
               /note="o346; 99 pct identical to BIOB_ECOLI SW: P12996"
               /codon_start=1
               /transl_table=11
               /product="biotin synthesis, sulfur insertion?"
               /protein_id="AAC73862.1"
               /db_xref="GI:1786992"
               /translation="MAHRPRTLSQVTELEFEKPLDILFEAQVYRQHFDEYOVST
               LSTRTGACPECKYCPOSSRYKTLGLEERLMEYVLESARKKAASSTFGCAAN
               KNPHERMDPYLEOMVQYKAMGLEACMTLGTISESOAQRANAGIDYNNHLDPSER
               YGNITITRYQERLDTLEKVRDAGIKVCSGIGVSGEVRDAGILLQLANLPMPES
               VPIMLVYKGTPLADNDVDAPDFIRITAVARIMPTSVYRISAGREOMEGQAMC
               FMAGANSFEYCGKLLTTPNPERDKDLQFKRLGLNPQOTAVLAGDNEQOHLBQALMT
               PPTDEYVNAAL"
               4530..3684
               /gene="bioF"
               /note="b0776"
               4530..3684
               /gene="bioF"
               /EC_number="2.3.1.47"
               /function="enzyme; Biosynthesis of cofactors, carriers:
               Biotin"
               /note="o384; 100 pct identical to BIOF_ECOLI SW: P12998"
               /codon_start=1
               /transl_table=11
               /product="8-amino-7-oxononanoate synthase"
               /protein_id="AAC73863.1"
               /db_xref="GI:1786993"
               /translation="MSWQKTIINAALDARRADALRRRYPAQAGRWLVADROYLNF
               SSNDYLGISHPQIIRAWQOGAEQFGIGSGSGHVSQSVYVQALBELAELEWLGYSFA
               LFIISGEANQAVIAAMAKEDRIARLASHLSLEAASLSPSOLRRFANHDVYHLAR
               LLASPCQGOAVTEGVFSDGDSAPLAEIOQVYQHGNGMLMDVDAAGTGVIGGQGRG
               SCMLQKVPPELLVTFPGKFGVSAVALCSSTVADYLLQFARHLITYTSMPPAQOAL"
Query Match 100.0%; Score 25; DB 1; Length 11022;
Best local Similarity 100.0%; Pred. No. 0.28;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 gcatcgccagcatcgatcgatcgt 25
Db 3582 GCATCGCCAGCATTCGATCCTCGT 3606
RESULT 13
AE005258 13501 bp DNA BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
DEFINITION of 155
ACCESSION AE005258 AE005174

```

```

VERSION      AE005258.1  GI:12513751
KEYWORDS
SOURCE       Escherichia coli O157:H7 EDL933.
ORGANISM     Escherichia coli O157:H7 EDL933
REFERENCE    Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
AUTHORS      1 (bases 1 to 13501)
              Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grobbeck,E.U., Davis,N.W., Lim,A., Dimantola,E., Potamousis,K.,
              Apodaca,U., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
TITLE        Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL      Nature 409 (6819), 529-533 (2001)
MEDLINE     11206551
PUBMED      11206551
REFERENCE    2 (bases 1 to 13501)
              Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grobbeck,E.U., Davis,N.W., Lim,A., Dimantola,E., Potamousis,K.,
              Apodaca,U., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
DIRECT SUBMISSION
JOURNAL      Submitted (22-OCT-2000) Laboratory of Genetics, University of
              Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
SOURCE       Location/Qualifiers
              1..13501
              /organism="Escherichia coli O157:H7 EDL933"
              /strain="EDL933"
              /serotype="O157:H7"
              /db_xref="taxon:155864"
              /note="enterohaemorrhagic"
              <1..7576
              /note="O-island #36; Region of the EDL933 chromosome not
              homologous to E. coli K-12 MG1655; Cryptic prophage
              CP-933K; includes one copy of the 13 bp direct repeat that
              flanks the prophage"
              66..665
              /gene="10mK"
              /note="20981"
              66..665
              /gene="10mK"
              /function="putative membrane: Other or unknown (Phage or
              prophage Related)"
              /note="Residues 1 to 199 of 199 are 72.86 pct identical to
              residues 1 to 199 of 199 from Genpept 118 :
              g117532789|gb|AA63231.1|AF151091_2 (AF151091) 10m
              [prophage P-E1bA]"
              /codon_start=1
              /transl_table=11
              /product="putative outer membrane protein of prophage
              CP-933K"
              /protein_id="AA655137.1"
              /db_xref="GI:12513752"
              /translation="MKRVCAALISAACIAVSGVPAMASRHOSTLSNGYTHASTDAP
              SDDINGINAKRYEFTDGLITFSYANEDQKTHYSTRWHEHYVRRRWKSVAG
              PSVRVNEVFSAYAMAGVAYSRSVSGDYRVRDNKRKTHDVLTGSDDAKYSNTSLAW
              GAGVQENPTEFSVAVDAVEKXSGSDMRTDGIIVGVYKF"
              719..2041
              /gene="20982"
              719..2041
              /gene="20982"
              /function="putative structure: Structural component (Phage
              or Prophage Related)"
              /note="Residues 164 to 440 of 440 are 68.79 pct identical
              to residues 381 to 645 of 645 from Genpept 118 :
              g114584361gb|AAD25464.1|AF125520_59 (AF125520) putative
              tail fiber protein [Bacteriophage 933W]"
              /codon_start=1
              /transl_table=11
              /product="putative tail component of prophage CP-933K"

```

```

              /protein_id="AA655138.1"
              /db_xref="GI:12513753"
              /translation="MGMAVQISGVCLKDCAKPTQNCITQKARNSTXYVNVYASE
              NPDEAGRISMDVEYGOYSVTLLEGGPPPSHAQTISVYEDSQGTINDLGLAMTEEDAR
              PEALRRPEQMEVEEARHAEKAKNAGEATTSARNAGISAKSAKASANAADTSADSEAR
              SAROAESAASKSEASASSASAEAKQASSELISQATSAELSKTASAGNARDA
              TTSTEKARSASASQAEQSRIAEDAVVARIPTVYVGPPEKGPAPDQGGKAGE
              KDDTPGAGATGRRGPGDGPAGKDBREKRETCGTGACGCGKDPGAGCA
              GQDGPKEGTGAAGPVGATGPGPKGPGGTQIRFRIGPKRILLETNSYGFPETDGALI
              TGTTFEDPPDAVQVQGMFQHILQVRKEDGPMQVKGIDVSGSDTGRTE"
              2004..2312
              /gene="20984"
              2004..2312
              /gene="20984"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"
              /note="Residues 1 to 102 of 102 are 98.03 pct identical to
              residues 28 to 129 of 129 from Genpept 118 :
              g114584371gb|AAD25465.1|AF125520_60 (AF125520)
              hypothetical protein [Bacteriophage 933W]"
              /codon_start=1
              /transl_table=11
              /product="unknown protein encoded by prophage CP-933K"
              /protein_id="AA655139.1"
              /db_xref="GI:12513754"
              /translation="MKMAVTOAELENDMNLKPKMQSLCGCKHDCENGSTPQLR
              LGPADLLESDENGITPEQDRVITQVYILDAKKQIQCVVRPLQILRADGTWENTGMR
              "
              2489..3469
              /gene="20985"
              2489..3469
              /gene="20985"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"
              /note="Residues 13 to 326 of 326 are 55.73 pct identical
              to residues 18 to 331 of 336 from Genpept 118 :
              g1169603671gb|AA63227.1 (AF170176) hypothetical protein
              predicted by Glimmer [Salmonella typhimurium LT7]"
              /codon_start=1
              /transl_table=11
              /product="unknown protein encoded by prophage CP-933K"
              /protein_id="AA655140.1"
              /db_xref="GI:12513755"
              /translation="MISPTRTFFHNSVNTVQSSPCQVTSFACKETELVIDEKPIFL
              QMPEPNPERYKKEVPIVNTKQHPYLDVNTNARIESDMIGTFVDSGFVNVQKTAFS
              KLERDEENWIIYREDVDPSMYDRKLSIDYHDIIICEORLKEDEKDEVLNLEKTAFS
              EISKQDSLISMVAKKRNHAWPDFRLLAKGEIFCTYVTKNHGISFGSGCIYLD
              MMILITGKGTIYAPDGISMHVDRRDSVNIENSALIVNSHNPALLGLSPMSHKVD
              AHPIYDGLGKGYKKYFNFTPLANNHPCDFIEFNHPIIMNTSQTCSW"
              3530..4522
              /gene="20986"
              3530..4522
              /gene="20986"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"
              /note="No significant matches"
              /codon_start=1
              /transl_table=11
              /product="unknown protein encoded by prophage CP-933K"
              /protein_id="AA655141.1"
              /db_xref="GI:12513756"
              /translation="MKIPSLQSNFNPSPACYSAPAIARNAENATADVLDIGKPIPL
              SAADLSNVYESVIRAVHDSSRLIDQHTVDMIGNTVLDALSSQTFRAVSGIHNK
              VHAIGCIKRYNEYLENEESSVKIDDITQSLTCLNGLYDVGOEIPICAGEGNDEPEY
              VSEVAPDIDSYSEMSQWQGLTHEITIIHHVTSDESDSGSNIIEIGPTEILARVQDELG
              MSVPDPRGYAEPEREAHRLRLNLRLNLAQAMHEENENAFPEPLGTTIDRYEASPDFT
              EYSAYSNIGYGFIIQGHDPFGGLAININDQANDIQLYHRAPIYITFGDVDKINQ"
              5350..6231
              /gene="20989"
              5350..6231
              /gene="20989"
              /function="orf; Other or unknown (Phage or Prophage
              Related)"

```

	gene	/note="No significant matches" /codon_start=1 /transl_table=1 /product="unknown protein encoded by prophage CP-933k" /protein_id="AAG55142.1" /db_xref="GI:12513757" /translation="MLSPYSVNLGSCSNNSITRLNLTSPDNRYLSSVPDAAVHSDGAQOV KVGRRYRVATDDNRKPCVTRESHSQCFTNLLHRIGMKRGESRKIEVMANASPVSAAM ERGVIRNRPDLDPVDAPPELPISVDYNRLTSPGNVIGKGNAVAYDAEADPATVLMK FHTSOSNEEYETSECRGNOXYGAGSARKITGNNDIIGIRDKINGSLINISTSPAO AEHAITYMPDRLECKGLFVDTTETNVLTRAKNEFPIDISSTNVSDRSMSQIMQ SYHGKGDLSIVLSLK1"
	CDS	6462..7160 /gene="20990" 6462..7160 /gene="20990" /function="orf; Other or unknown (Phage or Prophage related)" /note="No significant matches" /codon_start=1 /transl_table=1 /product="unknown protein encoded by prophage CP-933k" /protein_id="AAG55143.1" /db_xref="GI:12513758"
	gene	/translation="MRPTSLNVLVHGSRSSMSKDIDESLYKASSVQWIKNNPOLRFR QGHDNIYQQILEALDKIGSETGRVLLNLESISRKSETVYLHLNSSRIGVAHBD IDENHGRTSGDFECNCINAAYEPCGEISVYDFPAITVFHLLFLPNLINERLKVSSE SRASEQXSPILLTEARTVCGIGAFSEEVLSSENKFHEETGMRTSPYXSALIHHDDNT VSGFOOVRLLPIL1"
	CDS	complement(7668..8144) /gene="ybhb" /note="20992" complement(7668..8144) /gene="ybhb" /function="orf; Unknown function" /note="Residues 1 to 158 of 158 are 99.36 pct identical to residues 1 to 158 of 158 from Escherichia coli K-12 strain MG1655: B0773"
	gene	/codon_start=1 /transl_table=1 /product="orf, hypothetical protein" /protein_id="AAG55144.1" /db_xref="GI:12513759" /translation="MKLISNDLRDGDKLPHRHVNGMGYGDSNTISPLAWMDVPAGTK SFVYTCDPDAPTGGMMHMYVLLPADTRLPLPGFGSGIYAMPDVLIQTTPDGKTIG YDGAAPPKGETHERIFLVHADVERIDVDEBAGAMGFNFHSLASITTAIFS" complement(8203..9492) /gene="b10a" /note="20993"
	CDS	complement(8203..9492)
	Query Match	100.0%; Score 25; DB 1; Length 13501;
	Best Local Similarity	100.0%; Pval. No. 0.28;
	Matches	25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy	1	gcattgccagcattcgactctcgt 25
Db	9668	GCATTGCCAGCATTTCGATCTCTGT 9692
RESULT	14	
LOCUS	AP002553	297816 bp DNA BCF 07-MAR-2001
DEFINITION	Escherichia coli O157:H7 DNA, complete genome, section 4/20.	
ACCESSION	AP002553 BA000007	
VERSION	AP002553.1 GI:13360211	
KEYWORDS		
SOURCE	Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)	
ORGANISM	Escherichia coli O157:H7	
REFERENCE	Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia. 1 (sites)	

TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS
Complete nucleotide sequence of the prophage ϕ 2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic <i>Escherichia coli</i> O157:H7 derived from the Sakai outbreak	JOURNAL	GENES GENET. SYST. 74 (5), 227-239 (1999)		Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurikawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic <i>Escherichia coli</i> O157:H7 Sakai strain and an <i>Escherichia coli</i> K-12 strain MG1655	JOURNAL	SYST. APPL. MICROBIOL. 23 (3), 315-324 (2000)		Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C. H., Kurikawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C. and Shinagawa, H.
Complete nucleotide sequence of the prophage ϕ 1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic <i>Escherichia coli</i> O157:H7 strain derived from the Sakai outbreak	JOURNAL	GENE 258 (1-2), 127-139 (2000)		Yokoyama, K., Makino, K., Ohnishi, M., Kurikawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsuka, E., Nakayama, K., Murata, T., Tanaka, M., Ito, T., Iida, T., Takami, H., Honda, T., Sasaki, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shida, T., Hattori, M. and Shinagawa, H.
Complete genome sequence of enterohemorrhagic <i>Escherichia coli</i> O157:H7 and genomic comparison with a laboratory strain K-12	JOURNAL	DNA RES. 8 (1), 11-22 (2001)		Hayashi, T., Makino, K., Ohnishi, M., Kurikawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsuka, E., Nakayama, K., Murata, T., Tanaka, M., Ito, T., Iida, T., Takami, H., Honda, T., Sasaki, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shida, T., Hattori, M. and Shinagawa, H.
5 (bases 1 to 297816)	JOURNAL	21156231		Ohnishi, M., Kurikawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.
Direct Submission	JOURNAL	Submitted (26-JUN-2000)		Ken Kurokawa, Osaka University, Genome Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: kene@gen-info.osaka-u.ac.jp)
URL: http://www.gen-info.osaka-u.ac.jp/	JOURNAL	Fax: 81-6-6879-2047		
genome project.	JOURNAL			
Location/Qualifiers	JOURNAL			
1..297816	JOURNAL			
/organism="Escherichia coli O157:H7"	JOURNAL			
/strain="O157:H7"	JOURNAL			
/sub_strain="K12 0509952"	JOURNAL			
/db_xref="taxon:83334"	JOURNAL			
79..1245	JOURNAL			
/gene="ECs0753"	JOURNAL			
79..1245	JOURNAL			
/gene="ECs0753"	JOURNAL			
79..1245	JOURNAL			
/note="similar to SUCC_ECOLI g111706948 percent identity 100 in 388 aa (Conserved in E.coli K-12)"	JOURNAL			
100 in 388 aa (Conserved in E.coli K-12)"	JOURNAL			
/codon_start=1	JOURNAL			
/transl_table=11	JOURNAL			
/evidence-not_experimental	JOURNAL			
/product="succinyl-CoA synthetase beta subunit"	JOURNAL			
/protein_id="BAB34176.1"	JOURNAL			
/db_xref="GI:13360212"	JOURNAL			
/db_xref="GI:13360212"	JOURNAL			
/translation="MNLEHYQAKQLFARYGLDPAPYGYACTTPREAEKATIGAGPVPVVKCVHAGRGKAGVYVNSKEDIDAFENMIGKRLVTVQTDAGQSPVNOILVEAAVLDIAELVIGAVDSSRRVYFMASTEGVEIKVAEERPHLTHKVALDPLGPMRYOGDIEALKEVLECKLIVOCFTKTFMGATLFLERDLATLEIPLVITKGLDLCRKLKGLMGMDIYDGNLFPQDILREMDROSDPREDAQAOAMELWYALGNCIGVNGAGLGMGMMDIYKLHGEPANFLDNGGCACTKRYRYEARKIILSDKYATVYVFEGVITCDIILAGIIGAVAEAGVAVPVPVRLGNNNELDAKRLAOSGLNITLAKGLVDAAQVVAAVEGR"	JOURNAL			


```

gene
1245..2114
/ gene="ECs0754"
/ gene="ECs0754"
/ note="similar to SUCD_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSLIDKNTKVICGSGFTSGGTFHSEDAIAYGKMYGVTGPGK
GTHLGIPTVNTVREAVATGATATATATATATATATATATATATATATATATATAT
LDMITVVKLDEAGVNMIGPCGVITPGECKIGIOPGHINPKGVIGVRSGLTYTE
AVKOTDYGROSTCVGIGSDPIPGSNFIDILEMFEKDPQTEAIVMIGELGSGAEEA
AAVIREHVTKPVVGIAGVTAPKGRMGHAGAIAGGTADKFALEAAGVTVRS
LADIGELKTYLK"
2781..3686
/ gene="ECs0755"
/ gene="ECs0755"
/ note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g114170431sp|P32064|GCV_A_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELLVTEVVARRHESYTRAEELATQSAVEROVSAL
EEPLRTPLFSHKKRIPLNDGKYLGIVKTELNKLEEDTNTIMTQPTVOVIELAVN
PPTSHMLIPMLHEFTKLHPDIIVNHSLANGFPLNREYDAVIMRENFCAPMAEVEY
LPEEELIPVCGSILAMSDOKLSVAELITTEPLIHOSRITGMEWEMFALSGVSSPLVN
NGPRDLISMLIAVRSNLYVALLPRPAIQHDLSDGMVIPCDVIPRTGNFIMTQOE
EKSDSPHLQOFREWLKASVVDQM"
complement(3720..4322)
/ gene="ECs0756"
/ gene="ECs0756"
/ complement(3720..4322)
/ gene="ECs0756"
/ note="probable cob(1)alamin adenosyltransferase, similar
to cob(1)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
g11151481sp|P13040|BTUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative cob(1)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISFTHVROOROKLKEQVDFRVAATEKKGLIVFTGNGK
KSTAFGVTVRAVGHGTVGAQYIKGQMDGCEVLLDPLGVEFHIMTGTWETQNR
QADIDAEVWSESKRMILADKRDLDVLDDELPTLATHYLDTEEVIASLQNRPAQSY
IVTGRGCHSQLIKMADTVSEIRPVHAFDNGIOAPGIDW"
complement(4332..5984)
/ gene="ECs0757"
/ gene="ECs0757"
/ complement(4332..5984)
/ gene="ECs0757"
/ note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarate) - Escherichia coli
g11205981sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"

```

```

/translation="MSKPTFMOEPFLQNNKDGTEYLLISDQHTVTELDGEFVIRIAPE
ALTLARQAESEAFPLRSAGHLOQVASTINDQASNDKAYVALDLRNAAVSAGVLP
NCODTCTATIVASKGQIWTGNGDAEALSGISYFQSNRLRFSQNAFLDMYETVNTQ
TNPAQIDISAVAGDEYHFLCYNKGSGSNKAAKAOERKSLLOPEKLTAPLIEKKSL
GTACPPYHIAVVGGLSDQYLLKTLAKLSTYTNLPLSGDEGQAFRDILEKVL
EASQCPGIAQGGKGFADHIVIRLPRGGSCPLAMLSGADNRINAKINKHITWL
EKLEHNGOYIIPASLEENHQAHOVDLNRPLDMODLARLPVETRSLSGPIYVAR
DIARHAKRGLDSGEMPEYLNKHIIYVYAGPAKTPENMACSLGFTGSGMDGYIDTF
QAAGSTVMSLKSNGSCQVTDACCHKGNGGIGGAAALAOEYVSLRCLFEPBELG
MEAVMMMEVNIIPAFILVDKGNPFSDPEQHRCASCPAGH"
complement(6092..7372)
/ gene="ECs0758"
/ gene="ECs0758"
/ complement(6092..7372)
/ gene="ECs0758"
/ note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g11214671sp|P24943|GLT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLTWMLALVIGMIGVLTNTASPEAKLYAOEISIFT
TIFRLIKMLIAPLYSTLVYGIARKGDAKALGRFSKTLPTFCASLSILGILTV
NFPFGTGINFPAHGAETTGVAAPETTKVFISNAPFSTYDANAHNIIQIVESI
FLGCSLTAIGEGSAIVAHLDLSAHMLKLTGVMLFAVLVPAFVIALIARGLAVM
VSAIGTFMGEPFYFMLLMLVLLIGLAIVYGPCTIRLTRALSEPALLAFTSSGAAP
GTLEKLEQFSGVSPKIASFVLPISFNLVSGMAVCSFAVFTAOACNHLISGEQITM
LILMLTSKMGAGVPRAVSVYIATLQNPINPAGLILMGVDFPLDMGRSATVNSN
AMGAAYVSRWEGHCEGCGRKAKRNESVALP"
complement(7533..7850)
/ gene="ECs0759"
/ gene="ECs0759"
/ complement(7533..7850)
/ gene="ECs0759"
/ note="unknown"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

```

Query Match 100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.25; Indels 0; Gaps 0;
Matches 25; Conservative 0; Mismatches 0;

Oy 1 gcatcgccagcatcgatcctcgt 25
|||||

Db 92400 GCATCGCCAGCATTTGCATCCTCGT 92424

RESULT 15
AF250776
LOCUS
DEFINITION
AF250776 5526 bp DNA BCT 31-JAN-2001
uncultured bacterium pCosHE2
uncultured bacterium pCosHE2
Bacteria; environmental samples.
KEYWORDS
ORGANISM
REFERENCE
AUTHORS
TITILE
JOURNAL
MEDLINE

20575196

PUBMED 11133432
 REFERENCE 2 (bases 1 to 5526)
 AUTHORS Entcheva, P., Liebl, W. and Streif, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
 FEATURES
 source
 1..5526
 /organism="uncultured bacterium pcoshe2"
 /db_xref="taxon:143797"
 /clone="pcoshe2"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(52..528)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical 17.1 kDa protein in modc-bioa
 intergenic region"
 /protein_id="AAG60577.1"
 /db_xref="GI:12620125"
 /translation="MKLISNDLRDGLPHRHVNGMGYDGNISPHLAMDPAGTK
 SPVYTCYDPAPPGSGMHWVNLPAIDRVLPFGSGGLVAMPDGVLOTDPFGAG
 YDGAAPKGETHRYIFTVALDIERIDVDEGASGAWGVFNHSHLASITAMS"
 complement(587..1876)
 /gene="bioa"
 complement(587..1876)
 /gene="bioa"
 /note="7,8-diaminopelargonic acid
 synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase Bioa"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDLAFDQRIHMYPTSMSPLPYPPVSAEGCELLSDGR
 LVDMSSWMAATHIGNHPQLNAAKSDIDAMSHVPGITTHAPALECRKIVAMPPOP
 LKCLFLADSGSAVEAVAKMALQYQAKGEARQRLTFENGYGDFGMSVSCDDNS
 MSLKMGYLPENLFAPOSRMDGEMDEDMGPAALMAHHHEIAVITPEIYOGAC
 GMRMYPEMLKRIKICDREGILLADLETATGFCGRKIFACEHAFIADPILCKAL
 TGGTMTLSATITTRREVAETISNGEACGFMHGT PKGNPLACAAANASLAIESGMQO
 OVADIEVQREOLAPARDAEMVADVAVIGAVETTHVNAALAKREFVEGVIRP
 EKLITLMPYITILPQOLQRLTAANRAVDETFQC"
 1963..3003
 /gene="biob"
 1963..3003
 /gene="biob"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase Biob"
 /protein_id="AAG60579.1"
 /db_xref="GI:12620127"
 /translation="MAHRPRLTSLQVTELEFEKPLDLLFEAQVVRHOFDRQOVST
 LLSITGACPECKCYCROSSRYKTGLEAEERLMEVQVLESARAKAAAGSTRGGAAM
 KRNERNDFLEOMVOGKADGLACMTLGTISEQARLANAGLDVYNNHNDLSPER
 IGNITLTRYOERLDITLERKVDAGIKVSGGIVGIGETVKRAGILLQLANIPPEP
 VPINLVKVGKPLADNDVDAFDEIRITVAARIIMPTSVYRLSAGRGROMBOQAMC
 FMAGANSIFYCGKLLTTPNEBEDKDLQFLRKLGIMPQQTAVLAGDNEQQLLEALMT
 PTDVYNAAL"
 3000..4154
 /gene="biob"
 3000..4154
 /gene="biob"
 /note="8-amino-7-oxononanoate synthase"
 /codon_start=1
 /transl_table=11
 /product="KAPA synthetase Biof"
 /protein_id="AAG60580.1"
 /db_xref="GI:12620128"
 /translation="MSWQEKINAALDARRADALRRRYVNOGAGRMVLVADROYLNF
 SSNDYLGSHHPQILRAWQAGQFQIGSGSGVSGVYHQALEFELEAMWIGYSRA
 LLFISFAANAVAIAMAKEDRIADRLSHSLFANASLPSQLRRFANDVTHLR
 LLASPCGQQLVTEGVFSMDGSAPLAEIQOVTQOINGMVLMDVADHGTGIGSGRG

SCMLQKVKPELLVTEGKFGVSGAANVCSSTVADYLIQFARHLITSMPPAQAL
 RASLQVITRDECDARREKVLSTIAHFRAGVDLPPTLADSCAIQPLIVGNSRALOL
 AKELQOGCWVAIRPVPAGTARLRLITTAHEMODIDRLLEVLHNG"
 4141..4896
 /gene="biob"
 4141..4896
 /gene="biob"
 /note="reaction step prior to synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=11
 /product="biotin biosynthesis protein Bioc"
 /protein_id="AAG60581.1"
 /db_xref="GI:12620129"
 /translation="MATVKKQAIAAAFGRALYERHADLQROSADYILAMPKRYT
 HYLDAGCGPMTTRHMRHQAQVATLSPMLYQAKQKADADHYTLGDIESTPLATA
 TFDLANSMLAYOMCNLSTALRELYRYSKGVAFITLVGSLPELHQWQVADERP
 HANRFLPDEIEQSLNGVHQHIOPTLTFMDALSMRSLKIGATLHGRDRPIL
 TRSQLRQLQAMPOQGRKPLTYHLFGLVARE"
 4883..5526
 /gene="biob"
 4883..5526
 /gene="biob"
 /note="DTP synthetase"
 /codon_start=1
 /transl_table=11
 /product="dethiobiotin synthetase Bioc"
 /protein_id="AAG60582.1"
 /db_xref="GI:12620130"
 /translation="MIVSKRYFVNGTDEYKATVASCALLQAKKAGYRTAGKPPAS
 GSEKTPGRLNSDALALQORNSLQDLDTATNPITFAPPTSPHITISGKRPISVVS
 SGLRLEQADWAVLVEAGGWFPTLSQTFADWVQEOPLVILVGVKGCINHAML
 TQAOIQRHGLTLAGVANDVTPPKRKAETITLIRMIAPPLIGETPWLAEENE"
 1274 a 1507 c 1567 g 1178 t

Query Match 96.0%; Score 24; DB 2; Length 5526;

Best Local Similarity 100.0%; Pred. No. 0.86; Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcgccagcattcgatccgt 25
 Db 2053 CATGCCAGCATTTGCGATCTCGT 2076

RESULT 16

AEO03860 10593 bp DNA BCT 06-JUL-2000
 LOCUS Xylella fastidiosa, section 6 of 229 of the complete genome.
 DEFINITION AEO03860 AEO03849
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Xylella fastidiosa.
 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 Xylella.

REFERENCE 1 (bases 1 to 10593)
 AUTHORS Silvestri, M.L., Siqueira, M.J., de Souza, A.A., de Souza, A.P.,
 Terenzi, M.F., Truffi, D., Tsai, S.M., Tsuchiko, M.H., Vallada, H., Van
 Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Zago, M.A., Zatz, M.,
 Meidanis, J., and Setubal, J.C.
 TITLE The genome sequence of the plant pathogen *Xylella fastidiosa*. The
 Xylella fastidiosa Consortium of the Organization for Nucleotide
 Sequencing and Analysis, Sao Paulo, Brazil
 JOURNAL Nature 406 (6792), 151-157 (2000)

MEDLINE 2 (bases 1 to 10593)
 REFERENCE Simpson, A.J.G., Reinach, F.C., Arruda, P., Abreu, F.A., Acencio, M.,
 Alvarenga, R., Alves, L.M.C., Arya, J.E., Baia, G.S., Baptista, C.S.,
 Barros, M.H., Bonaccorsi, E.D., Bordin, S., Boye, D.M., Britomes, M.R.S.,
 Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M.,
 Carier, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C.R.,
 Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E.,


```

gene
  complement(6494..7495)
  /gene="Xf0068"
  /complement(6494..7495)
  /note="similar to SP126601 (percent identity: 52 %/query alignment coverage: 86.8 %/subject alignment coverage: 99.7 %): identified by sequence similarity; putative; ORF located using Glimmer/BSFinder"
  /codon_start=1
  /transl_table=11
  /product="hydroxybenzoate octaprenyltransferase"
  /protein_id="AAE82881.1"
  /db_xref="GI:9104838"
  /translation="MTINQEHQYVNNHNSFPPLESDNSQSVSMAYREFTSAI
  TLAMHNRIDPYWKLARGDRVGFLLLPYVWALMLAAGVPPWTLCTVGTGL
  TRSAGCVINDYIDRWLDIPYERTIRPLVTGAVSPNALIMEATMLIAFGVLTINQ
  LVLLSVGLFLMTYPLKRYTLPOVYGLAFAFGCIPMAFAIIGKVPDQALIV
  ANILMTAYDWCANVRDDDKMGAKSTALFADIDLTVGVLYTLMLFTLCIVGLR
  ATLSHTYWSLILAAVALIGYQFIARRREPTACFRAMNNWGMTIFAGIALATTH"
  7506..7582
  /gene="trna-Arg-1"
  /note="found by tRNAscan"
  /product="trna-Arg"
  7506..7582
  /gene="trna-Arg-1"
  /complement(7517..7909)
  /gene="Xf0069"

Query Match
  Best Local Similarity 81.6%; Score 20.4; DB 1; Length 10593;
  Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcatcgccagatcgcctc 22
  ||||| ||||| ||||| |||||
Db 3297 GCATCGTCAGCATTTGCATCCT 3318

RESULT 17
AF250770 965 bp DNA BCT 31-JAN-2001
LOCUS
DEFINITION
  uncultured bacterium pCosHE1 DAPA-aminotransferase (bioA) and
  biotin synthase (biob) genes, partial cds.
ACCESSION
  AF250770.1 GI:12620104
VERSION
  AF250770.1
KEYWORDS
  uncultured bacterium pCosHE1.
  uncultured bacterium pCosHE1.
  Bacteria: environmental samples.
ORGANISM
  Bacteria: environmental samples.
REFERENCE
  1 (bases 1 to 965)
  Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streif,W.R.
  Direct cloning from enrichment cultures, a reliable strategy for
  isolation of complete operons and genes from microbial consortia
  Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
JOURNAL
  MEDLINE
  20575196
  11133432
REFERENCE
  2 (bases 1 to 965)
  Entcheva,P., Liebl,W. and Streif,W.R.
  Direct Submission
  Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
  Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
  source
  1..965
  /organism="uncultured bacterium pCosHE1"
  /db_xref="taxon:143796"
  /clone="pCosHE1"
  /note="unknown organism, cosmid clone derived from
  environmental consortium"
  complement(<1..438)
  /gene="bioA"
  complement(<1..438)
  /gene="bioA"
  /note="7,8-diaminopelargonic acid"
  CDS
  gene
  complement(<1..438)
  /gene="bioA"
  /note="7,8-diaminopelargonic acid"

```

```

gene
  synthetase-aminotransferase; similar to BioA"
  /codon_start=1
  /transl_table=11
  /product="DAPA-aminotransferase"
  /protein_id="AAG60563.1"
  /db_xref="GI:12620105"
  /translation="MTQDLAFDQOHIMPTSMTRPXPYPVAVASAHQRRRLVD
  GMSWMAAIVGHNPRLANAKAQIEQMSHYMFGITRHPAVDLCRRXVAMTFDLEC
  VLAQSGSVAXEVDENGASTAPLPADRPERYHXAFFIPXIQ"
  525..>965
  /gene="biob"
  525..>965
  /note="similar to Biob"
  /codon_start=1
  /transl_table=11
  /product="biotin synthase"
  /protein_id="AAG60564.1"
  /db_xref="GI:12620106"
  /translation="MHHAHTMSQVTELEKPPLELMEFAQVHQHDPDRHYQST
  LLSIKTGACPEDKVCPOSARYKTGLESERIMEVEQVLDARAKAKNAGSTRFCGAAM
  KKPHTVPCPLQEQMVYGVKMGLEALYDPRHARRAXSACLLA"
  230 a 256 c 285 g 180 t 14 others
BASE COUNT
  ORIGIN

Query Match
  Best Local Similarity 76.8%; Score 19.2; DB 2; Length 965;
  Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcatcgccagatcgcctc 24
  ||||| ||||| ||||| |||||
Db 614 GCATCGTCAGCATTCGATCCGCG 637

RESULT 18
AF250768 4000 bp DNA BCT 31-JAN-2001
LOCUS
DEFINITION
  uncultured bacterium pCosF51 BioA-like protein (biob) gene, partial
  cds; BioB-like protein (biob) and KAPA-synthetase (biob) genes,
  complete cds; and biotin biosynthesis protein Bioc (bioc) gene,
  partial cds.
ACCESSION
  AF250768
  AF250768.1 GI:12620095
VERSION
  AF250768.1
KEYWORDS
  uncultured bacterium pCosF51.
  uncultured bacterium pCosF51.
  Bacteria: environmental samples.
ORGANISM
  Bacteria: environmental samples.
REFERENCE
  1 (bases 1 to 4000)
  Entcheva,P., Liebl,W., Johann,A., Hartsch,T. and Streif,W.R.
  Direct cloning from enrichment cultures, a reliable strategy for
  isolation of complete operons and genes from microbial consortia
  Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
JOURNAL
  MEDLINE
  20575196
  11133432
REFERENCE
  2 (bases 1 to 4000)
  Entcheva,P., Liebl,W. and Streif,W.R.
  Direct Submission
  Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
  Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES
  source
  1..4000
  /organism="uncultured bacterium pCosF51"
  /db_xref="taxon:143795"
  /clone="pCosF51"
  /note="unknown organism, cosmid clone derived from
  environmental consortium"
  complement(<1..1177)
  /gene="bioA"
  complement(<1..1177)
  /gene="bioA"
  /codon_start=1
  /transl_table=11
  CDS
  gene
  complement(<1..1177)
  /gene="bioA"
  /note="7,8-diaminopelargonic acid"

```


* 3903 4002: gap of 100 bp
* 4003 4712: contig of 710 bp in length
* 4713 4812: gap of 100 bp
* 4813 5516: contig of 704 bp in length
* 5517 5616: gap of 100 bp
* 5617 6211: contig of 595 bp in length
* 6212 6311: gap of 100 bp
* 6312 6996: contig of 685 bp in length
* 6997 7096: gap of 100 bp
* 7097 7793: contig of 697 bp in length
* 7794 7893: gap of 100 bp
* 7894 8558: contig of 665 bp in length
* 8559 8658: gap of 100 bp
* 8659 9372: contig of 714 bp in length
* 9373 9472: gap of 100 bp
* 9473 10178: contig of 706 bp in length
* 10179 10278: gap of 100 bp
* 10279 10974: contig of 696 bp in length
* 10975 11074: gap of 100 bp
* 11075 11786: contig of 712 bp in length
* 11787 11886: gap of 100 bp
* 11887 12608: contig of 722 bp in length
* 12609 12708: gap of 100 bp
* 12709 13430: contig of 722 bp in length
* 13431 13530: gap of 100 bp
* 13531 14241: contig of 711 bp in length
* 14242 14341: gap of 100 bp
* 14342 15050: contig of 709 bp in length
* 15051 15150: gap of 100 bp
* 15151 15826: contig of 676 bp in length
* 15827 15926: gap of 100 bp
* 15927 16644: contig of 718 bp in length
* 16645 16744: gap of 100 bp
* 16745 17455: contig of 711 bp in length
* 17456 17555: gap of 100 bp
* 17556 18259: contig of 704 bp in length
* 18260 18359: gap of 100 bp
* 18360 19087: contig of 728 bp in length
* 19088 19187: gap of 100 bp
* 19188 19902: contig of 715 bp in length
* 19903 20002: gap of 100 bp
* 20003 20693: contig of 691 bp in length
* 20694 20793: gap of 100 bp
* 20794 21500: contig of 707 bp in length
* 21501 21600: gap of 100 bp
* 21601 22320: contig of 720 bp in length
* 22321 22420: gap of 100 bp
* 22421 23112: contig of 692 bp in length
* 23113 23212: gap of 100 bp
* 23213 23898: contig of 686 bp in length
* 23899 23998: gap of 100 bp
* 23999 24689: contig of 691 bp in length
* 24690 24789: gap of 100 bp
* 24790 25504: contig of 715 bp in length
* 25505 25604: gap of 100 bp
* 25605 26293: contig of 689 bp in length
* 26294 26393: gap of 100 bp
* 26394 27107: contig of 714 bp in length
* 27108 27207: gap of 100 bp
* 27208 27920: contig of 713 bp in length
* 27921 28020: gap of 100 bp
* 28021 28710: contig of 690 bp in length
* 28711 28810: gap of 100 bp
* 28811 29504: contig of 694 bp in length
* 29505 29604: gap of 100 bp
* 29605 30316: contig of 712 bp in length
* 30317 30416: gap of 100 bp
* 30417 31130: contig of 714 bp in length
* 31131 31230: gap of 100 bp
* 31231 31933: contig of 703 bp in length
* 31934 32033: gap of 100 bp
* 32034 32741: contig of 708 bp in length
* 32742 32841: gap of 100 bp

* 32842 33546: contig of 705 bp in length
* 33547 33646: gap of 100 bp
* 33647 34352: contig of 706 bp in length
* 34353 34452: gap of 100 bp
* 34453 35151: contig of 699 bp in length
* 35152 35251: gap of 100 bp
* 35252 35972: contig of 721 bp in length
* 35973 36072: gap of 100 bp
* 36073 36792: contig of 720 bp in length
* 36793 36892: gap of 100 bp
* 36893 37583: contig of 691 bp in length
* 37584 37683: gap of 100 bp
* 37684 38399: contig of 716 bp in length
* 38400 38499: gap of 100 bp
* 38500 39201: contig of 702 bp in length
* 39202 39301: gap of 100 bp
* 39302 40006: contig of 705 bp in length
* 40007 40106: gap of 100 bp
* 40107 40806: contig of 700 bp in length
* 40807 40906: gap of 100 bp
* 40907 41623: contig of 717 bp in length
* 41624 41723: gap of 100 bp
* 41724 42416: contig of 693 bp in length
* 42417 42516: gap of 100 bp
* 42517 43321: contig of 705 bp in length
* 43322 44028: contig of 707 bp in length
* 44029 44128: gap of 100 bp
* 44129 44857: contig of 729 bp in length
* 44858 44957: gap of 100 bp
* 44958 45633: contig of 678 bp in length
* 45636 45735: gap of 100 bp
* 45736 46452: contig of 717 bp in length
* 46453 46552: gap of 100 bp
* 46553 47260: contig of 708 bp in length
* 47261 47360: gap of 100 bp
* 47361 48069: contig of 709 bp in length
* 48070 48169: gap of 100 bp
* 48170 48868: contig of 699 bp in length
* 48869 48968: gap of 100 bp
* 48969 49666: contig of 698 bp in length
* 49667 49766: gap of 100 bp
* 49767 50470: contig of 704 bp in length
* 50471 50570: gap of 100 bp
* 50571 51265: contig of 695 bp in length
* 51266 51365: gap of 100 bp
* 51366 52040: contig of 675 bp in length
* 52041 52140: gap of 100 bp
* 52141 52853: contig of 713 bp in length
* 52854 52953: gap of 100 bp
* 52954 53667: contig of 714 bp in length
* 53668 53767: gap of 100 bp
* 53768 54480: contig of 713 bp in length
* 54481 54580: gap of 100 bp
* 54581 55296: contig of 716 bp in length
* 55297 55396: gap of 100 bp

Query Match 76.88; Score 19.2; DB 77; Length 62522;
Best Local Similarity 87.58; Pred. No. 1.5e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2 catccacgacattcgatccgtcgt 25
Db 26177 CATCCACGACATTCGATCCCTCT 26200

RESULT 20
AC090154/c
LOCUS AC090154
DEFINITION Homo sapiens chromosome 8 clone RP11-705c3 map 8, LOW-PASS SEQUENCE
SAMPLING AC090154
ACCESSION AC090154.1 GI:12957783

KEYWORDS HTG: HTGS, PHASEO.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 62522)
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
Britten, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 8, clone RP11-705C3
Unpublished
2 (bases 1 to 62522)
Britten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Boguslavskiy, L., Bouckgele, B., Brown, A.,
Cannata, J., Campopiano, A., Choedel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., DeArrelano, K., Dewar, K., Diaz, J., S.,
Dodge, S., Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Hearford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Karatas, A., LaRoque, K., Lamazares, R., Landers, T.,
Lehoczky, J., Levine, R., Liu, G., Maclean, C., Macdonald, P.,
Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrum, J., Menus, L., Mihova, T., Mlenka, V.,
Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,
Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,
Roy, A., Santos, R., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Souarez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J.,
Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viet, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J.,
Zemek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (17-FEB-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11757
Center clone name: 705_C3

* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
* 703 802: contig of 702 bp in length
* 803 1486: contig of 684 bp in length
* 1487 1586: gap of 100 bp
* 1587 2276: contig of 690 bp in length
* 2277 2376: gap of 100 bp
* 2377 3081: contig of 705 bp in length
* 3082 3181: gap of 100 bp
* 3182 3902: contig of 721 bp in length
* 3903 4002: gap of 100 bp
* 4003 4712: contig of 710 bp in length
* 4713 4812: gap of 100 bp
* 4813 5516: contig of 704 bp in length
* 5517 5616: gap of 100 bp
* 5617 6211: contig of 595 bp in length
* 6212 6311: gap of 100 bp

6312 6996: contig of 685 bp in length
* 6997 7096: gap of 100 bp
* 7097 7793: contig of 697 bp in length
* 7794 7893: gap of 100 bp
* 7894 8558: contig of 665 bp in length
* 8559 8658: gap of 100 bp
* 8659 9372: contig of 714 bp in length
* 9373 9472: gap of 100 bp
* 9473 10178: contig of 706 bp in length
* 10179 10278: gap of 100 bp
* 10279 10974: contig of 696 bp in length
* 10975 11074: gap of 100 bp
* 11075 11786: contig of 712 bp in length
* 11787 11886: gap of 100 bp
* 11887 12608: contig of 722 bp in length
* 12609 12708: gap of 100 bp
* 12709 13430: contig of 722 bp in length
* 13431 13530: gap of 100 bp
* 13531 14241: contig of 711 bp in length
* 14242 14341: gap of 100 bp
* 14342 15050: contig of 709 bp in length
* 15051 15150: gap of 100 bp
* 15151 15826: contig of 676 bp in length
* 15827 15926: gap of 100 bp
* 15927 16644: contig of 718 bp in length
* 16645 16744: gap of 100 bp
* 16745 17455: contig of 711 bp in length
* 17456 17555: gap of 100 bp
* 17556 18259: contig of 704 bp in length
* 18260 18359: gap of 100 bp
* 18360 19087: contig of 728 bp in length
* 19088 19187: gap of 100 bp
* 19188 19902: contig of 715 bp in length
* 19903 20002: gap of 100 bp
* 20003 20693: contig of 691 bp in length
* 20694 20793: gap of 100 bp
* 20794 21500: contig of 707 bp in length
* 21501 21600: gap of 100 bp
* 21601 22320: contig of 720 bp in length
* 22321 22420: gap of 100 bp
* 22421 23112: contig of 692 bp in length
* 23113 23212: gap of 100 bp
* 23213 23898: contig of 686 bp in length
* 23899 23998: gap of 100 bp
* 23999 24689: contig of 691 bp in length
* 24690 24789: gap of 100 bp
* 24790 25504: contig of 715 bp in length
* 25505 25604: gap of 100 bp
* 25605 26293: contig of 689 bp in length
* 26294 26393: gap of 100 bp
* 26394 27107: contig of 714 bp in length
* 27108 27207: gap of 100 bp
* 27208 27920: contig of 713 bp in length
* 27921 28020: gap of 100 bp
* 28021 28710: contig of 690 bp in length
* 28711 28810: gap of 100 bp
* 28811 29504: contig of 694 bp in length
* 29505 29604: gap of 100 bp
* 29605 30316: contig of 712 bp in length
* 30317 30416: gap of 100 bp
* 30417 31130: contig of 714 bp in length
* 31131 31230: gap of 100 bp
* 31231 31933: contig of 703 bp in length
* 31934 32033: gap of 100 bp
* 32034 32741: contig of 708 bp in length
* 32742 32841: gap of 100 bp
* 32842 33546: contig of 705 bp in length
* 33547 33646: gap of 100 bp
* 33647 34352: contig of 706 bp in length
* 34353 34452: gap of 100 bp
* 34453 35151: contig of 699 bp in length
* 35152 35251: gap of 100 bp
* 35252 35972: contig of 721 bp in length

Query Match 76.8%; Score 19.2; DB 77; Length 62522;
 Best Local Similarity 87.5%; Pred. No. 1.5e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DB 61648 CATCACGACATCTGATCTCCTCT 61625

LOCUS AC016014 68419 bp DNA HTG 13-JUL-2000
 DEFINITION Homo sapiens clone RP11-26J10, LOW-PASS SEQUENCE SAMPLING.
 AC016014
 AC016014.2 GI:9134660
 VERSION HTG; HTGS_PHASB0.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 68419)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Homo sapiens, clone RP11-26J10

JOURNAL
REFERENCE
AUTHORS

COMMENT

TITLE
JOURNAL

Unpublished
 2 (bases 1 to 68419)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Batta, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
 Cooke, P., Deatellano, K., Dewar, K., Domingo, M., Donelan, L., Doyle, M.,
 Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gardyna, S., Grant, G., Hayes, B., Heaford, A., Horton, L.,
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lehoczy, J., Lien, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McQuirk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wymann, D., Ye, W.J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (18-NOV-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 13, 2000 this sequence version replaced gi:6449522.
 All repeats were identified using RepeatMasker.
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIPR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence.submissions@genome.wi.mit.edu
 Project Information
 Center project name: I4643
 Center clone name: 26_J10

NOTE: This record contains 76 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 789 888: gap of 100 bp in length
 889 1666: contig of 778 bp in length
 1667 1766: gap of 100 bp
 1767 2567: contig of 801 bp in length
 2568 2667: gap of 100 bp
 2668 3636: contig of 969 bp in length
 3637 3736: gap of 100 bp
 3737 4595: contig of 859 bp in length
 4596 4695: gap of 100 bp
 4696 5517: contig of 822 bp in length
 5518 5617: gap of 100 bp
 5618 6436: contig of 819 bp in length
 6437 6536: gap of 100 bp
 6537 7361: contig of 825 bp in length
 7362 7461: gap of 100 bp
 7462 8283: contig of 822 bp in length
 8284 8383: gap of 100 bp
 8384 9171: contig of 788 bp in length
 9172 9271: gap of 100 bp
 9272 10074: contig of 803 bp in length
 10075 10174: gap of 100 bp
 10175 10962: contig of 788 bp in length
 10963 11062: gap of 100 bp
 11063 11845: contig of 783 bp in length
 11846 11945: gap of 100 bp
 11946 12758: contig of 813 bp in length
 12759 12858: gap of 100 bp
 12859 13619: contig of 761 bp in length
 13620 13719: gap of 100 bp


```

* 13720 14542: contig of 823 bp in length
* 14543 14642: gap of 100 bp
* 14643 15428: contig of 786 bp in length
* 15429 15528: gap of 100 bp
* 15529 16328: contig of 800 bp in length
* 16329 16428: gap of 100 bp
* 16429 17221: contig of 793 bp in length
* 17222 17321: gap of 100 bp
* 17322 18099: contig of 778 bp in length
* 18100 18199: gap of 100 bp
* 18200 18980: contig of 781 bp in length
* 18981 19080: gap of 100 bp
* 19081 19869: contig of 789 bp in length
* 19870 19969: gap of 100 bp
* 19970 20807: contig of 838 bp in length
* 20808 20907: gap of 100 bp
* 20908 21769: contig of 862 bp in length
* 21770 21869: gap of 100 bp
* 21870 22649: contig of 780 bp in length
* 22650 22749: gap of 100 bp
* 22750 23545: contig of 796 bp in length
* 23546 23645: gap of 100 bp
* 23646 24442: contig of 797 bp in length
* 24443 24542: gap of 100 bp
* 24543 25345: contig of 803 bp in length
* 25346 25445: gap of 100 bp
* 25446 26227: contig of 782 bp in length
* 26228 26327: gap of 100 bp
* 26328 27126: contig of 799 bp in length
* 27127 27226: gap of 100 bp
* 27227 28039: contig of 813 bp in length
* 28040 28139: gap of 100 bp
* 28140 28965: contig of 826 bp in length
* 28966 29065: gap of 100 bp
* 29066 29854: contig of 789 bp in length
* 29855 29954: gap of 100 bp
* 29955 30736: contig of 782 bp in length
* 30737 30836: gap of 100 bp
* 30837 31593: contig of 757 bp in length
* 31594 31693: gap of 100 bp
* 31694 32483: contig of 790 bp in length
* 32484 32583: gap of 100 bp
* 32584 33391: contig of 808 bp in length
* 33392 33491: gap of 100 bp
* 33492 34285: contig of 794 bp in length
* 34286 34385: gap of 100 bp
* 34386 35223: contig of 838 bp in length
* 35224 35323: gap of 100 bp
* 35324 36128: contig of 805 bp in length
* 36129 36228: gap of 100 bp
* 36229 37039: contig of 811 bp in length
* 37040 37139: gap of 100 bp
* 37140 37899: contig of 760 bp in length
* 37900 37999: gap of 100 bp
* 38000 38779: contig of 780 bp in length
* 38780 38879: gap of 100 bp
* 38880 39673: contig of 794 bp in length
* 39674 39773: gap of 100 bp
* 39774 40594: contig of 821 bp in length
* 40595 40694: gap of 100 bp
* 40695 41475: contig of 781 bp in length
* 41476 41575: gap of 100 bp
* 41576 42363: contig of 788 bp in length
* 42364 42463: gap of 100 bp
* 42464 43252: contig of 789 bp in length
* 43253 43352: gap of 100 bp
* 43353 44179: contig of 827 bp in length
* 44180 44279: gap of 100 bp
* 44280 45104: contig of 825 bp in length
* 45105 45204: gap of 100 bp
* 45205 46034: contig of 830 bp in length
* 46035 46134: gap of 100 bp
* 46135 46936: contig of 832 bp in length

```

```

Qy 1 gcatccgacattcgatctcg 24
Db 28757 GCATCCGACCATTCGATCCAGC 28734

```

```

Query Match 76.8%; Score 19.2; DB 64; Length 68419;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

* 46937 47036: gap of 100 bp
* 47037 47869: contig of 833 bp in length
* 47870 47969: gap of 100 bp
* 47970 48770: contig of 801 bp in length
* 48771 48870: gap of 100 bp
* 48871 49668: contig of 798 bp in length
* 49669 49768: gap of 100 bp
* 49769 50560: contig of 792 bp in length
* 50561 50660: gap of 100 bp
* 50661 51418: contig of 758 bp in length
* 51419 51518: gap of 100 bp
* 51519 52301: contig of 783 bp in length
* 52302 52401: gap of 100 bp
* 52402 53189: contig of 788 bp in length
* 53190 53289: gap of 100 bp
* 53290 54081: contig of 792 bp in length
* 54082 54181: gap of 100 bp
* 54182 54984: contig of 803 bp in length
* 54985 55084: gap of 100 bp
* 55085 55863: contig of 779 bp in length
* 55864 55963: gap of 100 bp
* 55964 56778: contig of 815 bp in length
* 56779 56878: gap of 100 bp
* 56879 57657: contig of 779 bp in length
* 57658 57757: gap of 100 bp
* 57758 58554: contig of 797 bp in length
* 58555 58654: gap of 100 bp
* 58655 59472: contig of 818 bp in length
* 59473 59572: gap of 100 bp
* 59573 60360: contig of 788 bp in length
* 60361 60460: gap of 100 bp
* 60461 61240: contig of 780 bp in length
* 61241 61340: gap of 100 bp
* 61341 62195: contig of 855 bp in length
* 62196 62295: gap of 100 bp
* 62296 63086: contig of 791 bp in length
* 63087 63186: gap of 100 bp
* 63187 63978: contig of 792 bp in length
* 63979 64078: gap of 100 bp
* 64079 64874: contig of 796 bp in length
* 64875 64974: gap of 100 bp
* 64975 65754: contig of 780 bp in length

```

```

RESULT 22
AC084172
LOCUS Homo sapiens chromosome 8 clone CTD-2577H2 map 8, WORKING DRAFT
DEFINITION AC084172.2 CT:12658056
ACCESSION AC084172.2
VERSION AC084172.2
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 116334)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone CTD-2577H2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 116334)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abrahams,H., Allen,N.,
Anderson,S., Batta,N., Bastien,V., Beda,F., Boguslavsky,L.,
Boukhalter,R., Brown,A., Burkett,G., Campioano,A., Castle,A.,
Choepey,T., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

```

```

DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,
FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,
Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L.,
Iliev, I., Johnson, R., Jones, C., Kann, L., Kartas, A., Larroque, K.,
Lamazares, R., Landers, T., Lebeckzy, J., Levine, R., Lieu, C., Liu, G.,
Macdonald, P., Margolis, N., McCarthy, M., McEwan, P., McKernan, K.,
McNeeters, R., Melrim, J., Meneus, L., Mihova, T., Mienge, V.,
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zaimoun, J.,
Zimmer, A. and Zody, M.

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: L1377
Center clone name: 2577_H_2

----- Summary Statistics -----
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 11357 bases at least Q40
Consensus quality: 114958 bases at least Q30
Consensus quality: 115443 bases at least Q20
Insert size: 11500; agarose-fp
Insert size: 115734; sum-of-contrigs
Quality coverage: 7.3 in Q20 bases; agarose-fp
Quality coverage: 7.2 in Q20 bases; sum-of-contrigs

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contrigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
-----
1 1008: contrig of 1008 bp in length
* 1009 1108: gap of 100 bp
* 1109 5806: contrig of 4698 bp in length
* 5807 5906: gap of 100 bp
* 5907 11021: contrig of 5115 bp in length
* 11022 11121: gap of 100 bp
* 11122 16878: contrig of 5757 bp in length
* 16879 16978: gap of 100 bp
* 16979 24688: contrig of 7710 bp in length
* 24689 24788: gap of 100 bp
* 24789 56381: contrig of 31593 bp in length
* 56382 56481: gap of 100 bp
* 56482 116334: contrig of 59853 bp in length.
Location/Qualifiers
1. 116334
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="CTD-2577H2"
/clone_lib="CTD Human BAC"
misc_feature
1. 1008

```

```

/note="assembly-fragment"
1109. 5806
/note="assembly-fragment"
misc_feature
5907. 11021
/note="assembly-fragment"
misc_feature
11122. 16878
/note="assembly-fragment"
misc_feature
16979. 24688
/note="assembly-fragment"
misc_feature
24789. 56381
/note="assembly-fragment"
misc_feature
56482. 116334
/note="assembly-fragment"
BASE COUNT 33562 a 23964 c 24393 g 33811 t 604 others
ORIGIN
Query Match 76.8%; Score 19.2; DB 76; Length 116334;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
OY 2 catgcgcagcattcgcctcgt 25
Db 102265 CATCACGACATCTCATCTCCT 102288

```

```

RESULT 23
AC026294/C
LOCUS
DEFINITION
AC026294 186616 bp DNA HTG 01-SEP-2000
15 unordered pieces.
AC026294.3 GI:9958420
VERSION
AC026294.3 GI:9958420
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 186616)
Waterston, R.H.
The sequence of Homo sapiens clone
unpublished
2 (bases 1 to 186616)
Waterston, R.H.
Direct Submission
Submitted (21-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:8099313.
COMMENT
Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H.NH0705C03
----- Summary Statistics -----
Sequencing vector: Plasmid; 0%
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 177884 bases at least Q40
Consensus quality: 180695 bases at least Q30
Consensus quality: 182282 bases at least Q20
Insert size: 199000; agarose-fp
Insert size: 185216; sum-of-contrigs
Quality coverage: 4.02 in Q20 bases; agarose-fp
Quality coverage: 4.35 in Q20 bases; sum-of-contrigs
----- NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contrigs. The true order of the pieces
* is not known and their order in this sequence record is

```

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 2457: contig of 2457 bp in length
* 2458 2557: gap of unknown length
* 2558 6164: contig of 3607 bp in length
* 6165 6264: gap of unknown length
* 6265 8851: contig of 2587 bp in length
* 8852 12910: contig of 3959 bp in length
* 12911 13010: gap of unknown length
* 13011 17550: contig of 4540 bp in length
* 17551 17651: gap of unknown length
* 17652 23522: contig of 5872 bp in length
* 23523 23622: gap of unknown length
* 23623 28616: contig of 4994 bp in length
* 28617 28716: gap of unknown length
* 28717 39183: contig of 10467 bp in length
* 39184 39283: gap of unknown length
* 39284 47100: contig of 7817 bp in length
* 47101 47200: gap of unknown length
* 47201 59667: contig of 12467 bp in length
* 59668 59767: gap of unknown length
* 59768 73204: contig of 13437 bp in length
* 73205 73305: gap of unknown length
* 73306 87755: contig of 14451 bp in length
* 87756 87855: gap of unknown length
* 87856 105521: contig of 17666 bp in length
* 105522 138110: gap of unknown length
* 138111 138210: contig of 32489 bp in length
* 138211 186616: contig of 48406 bp in length.
  
```

FEATURES

```

source
1. 186616
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="8"
   /clone="RP11-705C3"
misc_feature
1. 2457
   /note="assembly_name:Contig30"
misc_feature
2558. 6164
   /note="assembly_name:Contig31"
misc_feature
6265. 8851
   /note="assembly_name:Contig32"
misc_feature
8952. 12910
   /note="assembly_name:Contig33"
misc_feature
13011. 17550
   /note="assembly_name:Contig34"
misc_feature
17651. 23522
   /note="assembly_name:Contig35"
misc_feature
23623. 28616
   /note="assembly_name:Contig36"
misc_feature
28717. 39183
   /note="assembly_name:Contig38"
misc_feature
39284. 47100
   /note="assembly_name:Contig37"
misc_feature
47201. 59667
   /note="assembly_name:Contig39"
misc_feature
59768. 73204
   /note="assembly_name:Contig40"
misc_feature
73305. 87755
   /note="assembly_name:Contig41"
misc_feature
87856. 105521
   /note="assembly_name:Contig42"
misc_feature
105622. 138110
   /note="assembly_name:Contig43"
misc_feature
138211. 186616
  
```

```

BASE COUNT      53026 a 36754 c 37995 g 57255 t 1586 others
ORIGIN
  
```

```

Query Match      76.8%; Score 19.2; DB 70; Length 186616;
Best Local Similarity 87.5%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
  
```

```

QY 2 catgcacagcattcgatccctgt 25
Db 43327 CATCACACGATCTCGATCTCCT 43304
  
```

```

RESULT 24
AC011276/c
LOCUS
DEFINITION
  Homo sapiens clone RP11-13J17, WORKING DRAFT SEQUENCE, 7 unordered
  pieces.
AC011276 80954 bp DNA HTG 06-DEC-2000
AC011276
  
```

```

AC011276.4 GI:11560229
VERSION
KEYWORDS
SOURCE
  HTG: HTGS_PHASE1; HTGS_DRAFT.
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
  
```

```

REFERENCE
1 (bases 1 to 80954)
  Birren, B., Linton, L., Nusbaum, C. and Lander, E.
  Unpublished
  Homo sapiens chromosome, clone RP11-13J17
  2 (bases 1 to 80954)
  Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
  Baldwin, J., Barna, N., Beckert, R., Boguski, M., Collins, S., Collymore, A.,
  Brown, A., Castle, A., Colangelo, R., Dewar, K., Domino, M., Donnellan, L., Doyle, M.,
  Cooke, P., DeBartolomeo, K., Dewar, K., Domino, M., Donnellan, L., Doyle, M.,
  Ferreira, P., Fitzhugh, W., Forrest, C., Funk, R., Gage, D.,
  Galagan, J., Gardy, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
  Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
  Lechoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
  McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., McDowell, J.,
  Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
  Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
  Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
  Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
  Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
  Direct Submission
  Submitted (05-OCT-1999) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  On Dec 6, 2000 this sequence version replaced gi:1730084.
  All repeats were identified using RepeatMasker:
  Smit, A.F.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  
```

```

TITLE
JOURNAL
COMMENT
  
```

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

Project Information

Center project name: L3339

Center clone name: 13-J-17

Assembly program: Phrap; version 0.960731

Consensus quality: bases at least Q40

Consensus quality: bases at least Q30

Consensus quality: bases at least Q20

Insert size: 170000; agarose-1p

Quality coverage: 0.0 in Q20 bases; sum-of-contigs

Quality coverage: 0.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

* 19977 20778: contig of 802 bp in length
 * 20779 20878: gap of 100 bp
 * 20879 21692: contig of 814 bp in length
 * 21693 21792: gap of 100 bp
 * 21793 22622: contig of 830 bp in length
 * 22623 22722: gap of 100 bp
 * 22723 23542: contig of 820 bp in length
 * 23543 23642: gap of 100 bp
 * 23643 24477: contig of 835 bp in length
 * 24478 24577: gap of 100 bp
 * 24578 25389: contig of 812 bp in length
 * 25390 25489: gap of 100 bp
 * 25490 26306: contig of 817 bp in length
 * 26307 26406: gap of 100 bp
 * 26407 27210: contig of 804 bp in length
 * 27211 27310: gap of 100 bp
 * 27311 28129: contig of 819 bp in length
 * 28130 28229: gap of 100 bp
 * 28230 29038: contig of 809 bp in length
 * 29039 29138: gap of 100 bp
 * 29139 29723: contig of 585 bp in length
 * 29724 29823: gap of 100 bp
 * 29824 30631: contig of 808 bp in length
 * 30632 30731: gap of 100 bp
 * 30732 31532: contig of 801 bp in length
 * 31533 31632: gap of 100 bp
 * 31633 32447: contig of 815 bp in length
 * 32448 32547: gap of 100 bp
 * 32548 33347: contig of 800 bp in length
 * 33348 33447: gap of 100 bp
 * 33448 34246: contig of 799 bp in length
 * 34247 34346: gap of 100 bp
 * 34347 35150: contig of 804 bp in length
 * 35151 35250: gap of 100 bp
 * 35251 36035: contig of 785 bp in length
 * 36036 36135: gap of 100 bp
 * 36136 36941: contig of 806 bp in length
 * 36942 37041: gap of 100 bp
 * 37042 37859: contig of 818 bp in length
 * 37860 37959: gap of 100 bp
 * 37960 38778: contig of 819 bp in length
 * 38779 38878: gap of 100 bp
 * 38879 39686: contig of 808 bp in length
 * 39687 39786: gap of 100 bp
 * 39787 40616: contig of 830 bp in length
 * 40617 40716: gap of 100 bp
 * 40717 41545: contig of 829 bp in length
 * 41546 41645: gap of 100 bp
 * 41646 42457: contig of 812 bp in length
 * 42458 42557: gap of 100 bp
 * 42558 43379: contig of 822 bp in length
 * 43380 43479: gap of 100 bp
 * 43480 44279: contig of 800 bp in length
 * 44280 44379: gap of 100 bp
 * 44380 45193: contig of 814 bp in length
 * 45194 45293: gap of 100 bp
 * 45294 46089: contig of 796 bp in length
 * 46090 46189: gap of 100 bp
 * 46190 46999: contig of 810 bp in length
 * 47000 47099: gap of 100 bp
 * 47100 47914: contig of 815 bp in length
 * 47915 48014: gap of 100 bp
 * 48015 48820: contig of 806 bp in length
 * 48821 48920: gap of 100 bp
 * 48921 49629: contig of 709 bp in length
 * 49630 49729: gap of 100 bp
 * 49730 50532: contig of 803 bp in length
 * 50533 50632: gap of 100 bp
 * 50633 51445: contig of 813 bp in length
 * 51446 51545: gap of 100 bp
 * 51546 52345: contig of 800 bp in length
 * 52346 52445: gap of 100 bp
 * 52446 53346: contig of 801 bp in length

Query Match 74.4%; Score 18.6; DB 73; Length 112808;
 Best Local Similarity 84.0%; Pred. No. 2.7e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcatgcacgattcattccatcgt 25

Db 42421 GCATGCCACGATTGATCCCGT 42397

RESULT 26
 AC055855
 LOCUS
 DEFINITION
 Homo sapiens chromosome 15 clone RP11-352G18 map 15, WORKING DRAFT

ACCESSION
 AC055855
 VERSION
 AC055855.3 GI:8084722
 KEYWORDS
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 human.
 ORGANISM
 Homo sapiens

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Homo sapiens chromosome 15, clone RP11-352G18
 2 (bases 1 to 137599)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Baatien, V., Beda, F.,
 Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
 Campoliano, A., Castelle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heathford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
 Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienna, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, F., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testaye, S., Theodore, J., Tittrell, A., Travers, M., Trigglio, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE
 JOURNAL
 COMMENT
 Direct Submission
 Submitted (18-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 25, 2000 this sequence version replaced gi:1886628.
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: 352-G-18
 Center clone name: 352-G-18
 ----- Summary Statistics
 Sequencing vector: M13: M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 121247 bases at least Q40
 Consensus quality: 129329 bases at least Q30
 Consensus quality: 132685 bases at least Q20
 Insert size: 149000; agarose-fp
 Quality coverage: 3.5 in Q20 bases; agarose-fp
 Quality coverage: 3.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 28 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1      1180: contig of 1180 bp in length
*      1181 1280: gap of 100 bp
*      1281 2901: contig of 1621 bp in length
*      2902 3001: gap of 100 bp
*      3002 4102: contig of 1101 bp in length
*      4103 4202: gap of 100 bp
*      4203 5855: contig of 1653 bp in length
*      5856 5955: gap of 100 bp
*      5956 7517: contig of 1562 bp in length
*      7518 7617: gap of 100 bp
*      7618 8921: contig of 1304 bp in length
*      8922 9021: gap of 100 bp
*      9022 10045: contig of 1024 bp in length
*      10046 10145: gap of 100 bp
*      10146 12130: contig of 1985 bp in length
*      12131 12230: gap of 100 bp
*      12231 14274: contig of 2044 bp in length
*      14275 14374: gap of 100 bp
*      14375 15874: contig of 1500 bp in length
*      15875 15974: gap of 100 bp
*      15975 18366: contig of 2392 bp in length
*      18367 18466: gap of 100 bp
*      18467 21164: contig of 2698 bp in length
*      21165 21264: gap of 100 bp
*      21265 23595: contig of 2331 bp in length
*      23596 23695: gap of 100 bp
*      23696 25968: contig of 2273 bp in length
*      25969 26068: gap of 100 bp
*      26069 28594: contig of 2526 bp in length
*      28595 28694: gap of 100 bp
*      28695 30881: contig of 2187 bp in length
*      30882 30981: gap of 100 bp
*      30982 33464: contig of 2483 bp in length
*      33465 33564: gap of 100 bp
*      33565 37146: contig of 3582 bp in length
*      37147 37246: gap of 100 bp
*      37247 41197: contig of 3951 bp in length
*      41198 41297: gap of 100 bp
*      41298 44344: contig of 3047 bp in length
*      44345 44444: gap of 100 bp
*      44445 47926: contig of 3482 bp in length
*      47927 48026: gap of 100 bp
*      48027 53040: contig of 5014 bp in length

```

FEATURES
 source

```

*      53041 53140: gap of 100 bp
*      53141 60176: contig of 7036 bp in length
*      60177 60276: gap of 100 bp
*      60277 68162: contig of 7886 bp in length
*      68163 68262: gap of 100 bp
*      68263 75983: contig of 7721 bp in length
*      75984 76083: gap of 100 bp
*      76084 90034: contig of 13951 bp in length
*      90035 90134: gap of 100 bp
*      90135 111925: contig of 21791 bp in length
*      111926 112025: gap of 100 bp
*      112026 137599: contig of 25574 bp in length.
          Location/Qualifiers
            1. 137599
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /chromosome="15"
              /map="15"
              /clone="RP11-352G18"
              /clone_lib="RP11 Human Male BAC"
            1. 1180
              /note="assembly-fragment"
            1281. 2901
              /note="assembly-fragment"
            3002. 4102
              /note="assembly-fragment"
            4203. 5855
              /note="assembly-fragment"
            5956. 7517
              /note="assembly-fragment"
            7618. 8921
              /note="assembly-fragment"
            9022. 10045
              /note="assembly-fragment"
            10146. 12130
              /note="assembly-fragment"
            12231. 14274
              /note="assembly-fragment"
            14375. 15874
              /note="assembly-fragment"
            15975. 18366
              /note="assembly-fragment"
            18467. 21164
              /note="assembly-fragment"
            21265. 23595
              /note="assembly-fragment"
            23696. 25968
              /note="assembly-fragment"
            26069. 28594
              /note="assembly-fragment"
            28695. 30881
              /note="assembly-fragment"
            30982. 33464
              /note="assembly-fragment"
            33565. 37146
              /note="assembly-fragment"
            37247. 41197
              /note="assembly-fragment"
            41298. 44344
              /note="assembly-fragment"
            44445. 47926
              /note="assembly-fragment"
            48027. 53040
              /note="assembly-fragment"
            53141. 60176
              /note="assembly-fragment"
            60277. 68162
              /note="assembly-fragment"
            vector_side:right"
            68263. 75983
              /note="assembly-fragment"
            clone_end:SP6

```


misc_feature	/db_xref="GI:9244993"			
	/translation="MNINILPDILFVQIVDGSFSAVARONGITPESAVSRSVRLER EMGCKILQRTTRKRLKLSDAGETIYQIAQOMLEAAROMASAGSRQVAGCKITLISYPK AVGGEIVHLPMAMAFPHRYPOVDVCLRIEDLPDTIDGIDLALITDTPSGILGKPL MPLRIHYVCAATEAVLQOHGTPTYPQDILRAHSCISLGETPADARWRKFRREGTETVYQG RYANATVAVRLDAVRQHLGISTLPLEFARALANGIYOVLPEMEFTISSGDIWLILW AGCRHMPARMRAMIDYLTSETVPALNAGSTPEPAK"			
	3285..3344			
	/gene="ltra"			
	/note="encodes putative DNA-binding helix-turn-helix motif"			
BASE COUNT	995 a	1277 c	1145 g	887 t
ORIGIN				
Query Match	72.88;	Score 18.2;	DB 3;	Length 4304;
Best Local Similarity	87.08;	Pred. No. 4.7e+02;		
Matches	20; Conservative	0; Mismatches	3; Indels	0; Gaps 0;
Py	2 catcgccagatctgcacctcg 24			
Db	273 CATGCCAGCATGAGATCGTCG 251			
RESULT 31				
AE004457				
LOCUS	AE004457 10239 bp	DNA	BC1	30-AUG-2000
DEFINITION	Pseudomonas aeruginosa PA01, section 18 of 529 of the complete genome.			
ACCESSION	AE004457	AE004091		
VERSION	AE004457.1	GI:9946024		
KEYWORDS	Pseudomonas aeruginosa.			
SOURCE	Pseudomonas aeruginosa.			
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.			
REFERENCE	1 (bases 1 to 10239)			
AUTHORS	Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S., Huftnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H., Brody,L.L., Goltzy,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y., Folger,K.R., Kas,A., Larbig,K., Lim,R., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R., Smith,K., Spencer,D., Wong,G.K., Wu,Z., and Paulsen,I.T.			
TITLE	Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen			
JOURNAL	Nature 406 (6799), 959-964 (2000)			
MEDLINE	20437337			
REFERENCE	2 (bases 1 to 10239)			
AUTHORS	Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Huftnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H., Wong,G.K., Wu,Z., Paulsen,I.T., Reizer,J., Sater,M.H., Hancock,R.E.W., Lory,S. and Olson,M.V.			
TITLE	Direct Submission			
JOURNAL	Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA			
FEATURES	Location/Qualifiers			
source	1..10239			
	/organism="Pseudomonas aeruginosa"			
	/strain="PA01"			
	/db_xref="taxon:287"			
gene	155..880			
	/gene="PA0190"			
CDS	155..880			
	/gene="PA0190"			
	/codon_start=1			
	/transl_table=11			
	/product="probable acid phosphatase"			
	protein_id="AAG03580.1"			
	/db_xref="GI:9946025"			
	/translation="MNNKTLTCLPSLLICSLIAPISCGETAAAPYLAHPRLADLYLE			

[illegible]

```

gene
    /gene="pntA"
    /note="PA0195; PA0195 - This gene contains a frameshift
    that has been confirmed in the final assembly. Contains
    high homology to C-terminus of pntA"
    8602..10038
    /gene="pntB"
    /note="PA0196"
    8602..10038
    /gene="pntB"
    /codon_start=1
    /transl_table=11
    /product="pyridine nucleotide transhydrogenase, beta
    subunit"
    /protein_id="AAG03585.1"
    /db_xref="GI:9946030"
    /translation="MSMNLVTLILIASVCFIOALKLSHPTSRGNI.PGVGMIA
    VATTGLVFKEIATTTGGIVGLVGTAGTSMARKEMTKMPELVAFPMHMG
    LAEVEIAAVERPOSIGIVAHIGDTIPGTRLEIFGALIGATTPSGVIATGRKLSG
    KYKRFPGTPVPSGOHLINLVIGLTLGLIVFMTCNLIFAVMLAEVLGVLI
    IIPIGADMVYVSMNSISGMKAGIGFSLNNSMLIAGSLVSSGAILISYIMCKAM
    NRSFENVILGFGFAGADAGPAGSKGQRPVKSADADASELITNADSVITVPGYGLAV
    ARAQHALMELAEKLTIRGVTYKFAIHPVAGRMGHNMLLAEAEVPEYEQVEYEDINS
    EFGQTDVVLVGAENYVPAKNDPKSPIAGMPTLEAYKKAKTVIVNKRSAAGSYAGAD
    NELFYIDTKTMVFEGDAKKVIEDMYKAVE"

BASE COUNT      1705 a      3582 c      3290 g      1662 t
ORIGIN

Query Match      72.8%; Score 18.2; DB 1; Length 10239;
Best Local Similarity 87.0%; Pred. No. 4.6e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy      2 catcgccagcttcgcgtccg 24
      ||||| |||| ||||| |||||
Db      9861 CATGCCGCGCATGCCGATCCTCG 9883

RESULT 32
RLE310723      11382 bp      DNA      BCT      11-MAR-2001
LOCUS      Rhizobium leguminosarum tonB gene, hmu operon (hmuPV genes) and
            ipoz gene.
ACCESSION      AJ310723.1 GI:13277325
VERSION      hmuV uptake protein; HmuP gene; HmuS gene; HmuU gene;
KEYWORDS      hmuV gene; ORF1; ORFA; ORFB; ORPC; ORPD; ORPE; ipoz gene; sigma
            factor; tcrx gene; tcrY gene; tonB gene; two component regulator.
SOURCE      Rhizobium leguminosarum.
ORGANISM      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
            Rhizobiaceae; Rhizobium.
REFERENCE      1 (bases 1 to 11382)
AUTHORS      Wexler M., Yeoman K.H., Stevens J.B., De Luca N.G., Savers G. and
            Johnston A.W.B.
TITLE      Rhizobium leguminosarum tonB is required for siderophore and haem
            uptake and is next to hmu haem uptake genes
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 11382)
AUTHORS      Wexler M.
TITLE      Direct Submission
JOURNAL      Submitted (07-MAR-2001) Wexler M., Biological Sciences, University
            of East Anglia, Earlham Rd Norwich, Norfolk, NR4 7UJ, UNITED
            KINGDOM

FEATURES
    source      1..11382
                location/Qualifiers
                /organism="Rhizobium leguminosarum"
                /strain="8401(PRL11)"
                /sub_species="biovar viciae"
                /db_xref="taxon:384"
                /country="United Kingdom"
                /complement(1..315)
                /function="unknown"
                /note="ORFA"

CDS
    mat_peptide
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="CAC34386.1"
    /db_xref="GI:13277326"
    /translation="MYTAMNRFKVAATGSGDEFTVWRNRDSSLPEVPEFVRLRGK
    VNDEEGYTLSSHTYWKSEADFPQNMTKSMSFPAHRNMGDKHAIYKGGPVEGFNVVD
    GI"
    complement(4..312)
    /note="ORFA"
    /product="hypothetical protein"
    complement(325..597)
    /function="unknown"
    /note="ORFB"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="CAC34387.1"
    /db_xref="GI:13277327"
    /translation="MNDQRPDGFWMIGLHKLAAQTGEGLVPELYELFQOHAERQOIYQ
    NVMPPTWPARMPGEVATRPGLGPAANGDNVLAFPSSRAARAGKRKA"
    complement(328..594)
    /note="ORFG"
    /product="hypothetical protein"
    complement(594..1001)
    /function="unknown"
    /note="ORFC"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="CAC34388.1"
    /db_xref="GI:13277328"
    /translation="MARKGKGSNREVDSDAGDAGQASLGRSLKDGKSLFYVGRDC
    QVAHLROICSNFGAELIHHDGILREAVSRITVPSVDCVCPIDICSHDACLRVKG
    CKKFGKAFIPVNGSKSSLERALQTMNRDNR"
    complement(597..998)
    /note="ORFC"
    /product="hypothetical protein"
    complement(1109..2473)
    /gene="tonB"
    complement(1109..2473)
    /gene="tonB"
    /function="membrane protein; transduces energy to outer
    membrane receptors"
    /note="Size of TonB protein not experimentally determined;
    other possible start sites are at 2431 (440 residues) or
    2158 (349 residues)"
    /codon_start=1
    /transl_table=11
    /evidence-experimental
    /product="TonB protein"
    /protein_id="CAC34389.1"
    /db_xref="GI:13277329"
    /translation="MNPBGHGQTVRGIOMAISAKRSROYLIGEPDAGSLDNMNMHP
    GHEISLDLNVORPAGAEVVAHARPAQIPAPDPHEAPIASVPAPIDAAVEKOED
    KKPVRRLVLCIGSFPHAGLIALVAFMPTPNEAIEPAGAEVSVVYVGGSDVQDT
    SAGDEIQRORPEVASEVPEPDYVQSEAEFINATYTPPCAQPVETIEMETVQSOE
    VTRVSPREVAAPREVLVSESPASFAQPMATVAPQPMEDIAQATVPEVPTAVQ
    PTAAPREVAATVEVSPEDKPKRAKPKVEKKOSKRAKPAKAGEGSDREDSTRG
    TVNGOHGPDGTGTTGGTGDGMSAAVNPGRKIRRAVAVPDEYKNSKSGMTG
    RIQTLINTGTVASVARSSEIGIELDKAVLDGVRRAAPPPLPSEWKGSMFTQEV
    QVTR"
    complement(1112..2470)
    /gene="tonB"
    /product="TonB"
    complement(2487..2912)
    /function="unknown"
    /note="ORF1"
    orf1 is co-transcribed with tonB
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="CAC34390.1"

```

```

mat_peptide
  /db_xref="GI:13277330"
  /translation="MATVRFPAVAAGVYMAVSSVQADATAKLDVELNALASQK
  CMTVEAENNIQAPISKISFELAFENDKNAVDRITLDFRDLPGCKRRVORDMPNVK
  CEVTIRIITNDPVCDPATGCMGLVTRQISVPEFG"
  complement(2490..2909)
  /note="ORF1"
  /product="hypothetical protein"
  /gene="hmup"
  /gene="hmup"
  /gene="hmup"
  /function="involved in hemin uptake"
  /codon_start=1
  /transl_table=11
  /product="Hmup hemin uptake protein"
  /protein_id="CAC34391.1"
  /db_xref="GI:13277331"
  /translation="MYEKPDNFKHVPLOSEPAAGHRIVESADLFRGTNEIMLRHGLV
  YRKITRQCKLILNK"
  /gene="hmup"
  /gene="hmup"
  /product="Hmup hemin uptake protein"
  /gene="hmus"
  /gene="hmus"
  /function="putative hemin degrading protein"
  /codon_start=1
  /transl_table=11
  /evidence="experimental"
  /product="Hmus protein"
  /protein_id="CAC34392.1"
  /db_xref="GI:13277332"
  /translation="MTEQTRPAPAEIRTERAENPKRERDIAQLKISPAALVAETG
  ISVTRIDGALKLIERVAGIIGEMALSRNESAVHEKIGVPEMKSCQAALVGENID
  LRTPSRMEHFAVSKKDDOERLSQYFDKGNVAHKVHLRPSSEVAIYVAELK
  LEDOSQFEAEETNSAADDADYSDRELDMNSKLDTHTQFGMLKRLIGQAAVART
  VGDYAWRLDNSATAMHASVSKGLPICFPAVNGSIQIHSGPIFNQSMGPMINIM
  DPTFHLRLRODHTAETWAVAKPTTGHVTSLEAYNAEGEMITTFEKRQSGSDERAM
  REITEMI.PRAASVAA"
  /gene="hmus"
  /gene="hmus"
  /product="Hmus protein"
  /gene="hmus"
  /gene="hmus"
  /function="periplasmic hemin binding protein"
  /codon_start=1
  /transl_table=11
  /evidence="experimental"
  /product="Hmup protein"
  /protein_id="CAC34393.1"
  /db_xref="GI:13277333"
  /translation="MTMRNMLRRIRTRKAOLATAAVALPLIPTAPRIEGFAVRAAH
  ADKKLIDTSRLVSGDITREIVVAGESRLIARDTSMPEALIKGVNARALSP
  EGLIAMPRTAIIIVDSGPOEALTVLKNASVPREYPSATRRGCIITAKIDRVTLGV
  PDKAKALEEVAAADLDAIDAERKPEAKRVFLTISAQNGRIMASGTGTAADGVK
  LAGAKIRVGAPEGKPLTDEAIIEAKPDIIIMNRGSGAGCTKMEDLLAOPAVLTPAG
  EKRAIIFMDGVYLLGEGPRFAAARELSAIVG"
  /gene="hmup"
  /gene="hmup"
  /product="Hmup protein"
  /gene="hmus"
  /gene="hmus"
  /function="hemin permease"
  /codon_start=1
  /transl_table=11
  /evidence="experimental"
  /product="Hmup protein"
  /protein_id="CAC34394.1"

```

```

  /db_xref="GI:13277334"
  /translation="MALPQMAEERREPEPADIREIRQADRRRLALALAVASV
  FSMLEFSTTASDSDSTIDVISNMGSEFALSTDRITIIPIRLPAILIGLIGASLAV
  SGIVMGLFRPLDPCLVGVSSASGAYAMVYSGLAAPQALIGIYALPAAPAG
  GGIATYTLILYRIATRHGQTSVATMLGLIALGALALITGLITIMADQDLRTWS
  MSLAGATWTRIAAASP1IILSFPALPFRMAGLATALIGRAAFHMKVPQRIKNVI
  VGVAAATGASVAVSGIGIFVGIYVPHILRMAIGDHRFLPAAALLGSLILIRADVLA
  RTIYAPAEIRPILGITTAAGGPFILMLRQSRRLAL"
  /gene="hmup"
  /product="Hmup protein"
  /gene="hmus"
  /gene="hmus"
  /function="Atpase component of hemin transporter"
  /codon_start=1
  /transl_table=11
  /evidence="experimental"
  /product="Hmup protein"

```

Query Match 72.8%; Score 18.2; DB 3; Length 11382;
 Best Local Similarity 87.0%; Pred. No. 4.6e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcatgcagcatcgcaccc 23
 Db 10445 GCAGCGCGGATTCGATCCTC 10467

RESULT 33
 AE004541/C
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

AE004541 11840 bp DNA BCT 30-AUG-2000
 Pseudomonas aeruginosa PA01, section 102 of 529 of the complete genome.
 AE004541.1 GI:9947024
 Pseudomonas aeruginosa.
 Pseudomonas aeruginosa.
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.

REFERENCE
 AUTHORS
 1 (bases 1 to 11840)
 Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S., Huftagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltzy, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Laidig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K., Wu, Z. and Paulsen, I.T.
 Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
 Nature 406 (6799), 959-964 (2000)

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS

2 (bases 1 to 11840)
 Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P., Hickey, M.J., Brinkman, F.S., Huftagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltzy, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Laidig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K., Wu, Z., Paulsen, I.T., Reizer, J., Sater, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.
 Direct Submission
 Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA

TITLE
 JOURNAL

FEATURES
 source
 1..11840
 /organism="Pseudomonas aeruginosa"
 /strain="PA01"
 /db_xref="taxon:287"
 complement(130..1326)
 /gene="PA1107"
 complement(130..1326)
 /gene="PA1107"

```

/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG04496.1"
/db_xref="GI:9947025"
/translation="WARTLESIOHAIIEFNSSLAIPALLENISRLTHLALLTPFENLI
HLVSLKDFRDLGHOIMROIMYAHGAMALLFGIGILLMLRROPKIMRRLI
LLGGAGIIGFVYACIDORTSNITPLLACFACAMFTLIPAVAVPFGIAMLAF
VMAHQADPOLRLSNQANGIARGLILLIIMHGVRLROOKTELOROLEEK
NRQLEYLAGHDPIGTLENRRFDOLVIMELKRIAROPPLSLWDDHKEYINDRYG
HPGDEVEYRHAANLLRGTRSSDSYARLGGEEFILLPTVAEQOARSLAEKLRLEG
TPLATQDEVILCLTISFVASLDGGQASVEOLYSADQALYRAKSSGRNVEYTRIAS
LAESLR"
complement(1443..2600)
/gene="PA1108"
complement(1443..2600)
/gene="PA1108"
/codon_start=1
/transl_table=11
/product="probable MFS transporter"
/protein_id="AAG04497.1"
/db_xref="GI:9947026"
/translation="MPDRPRLIYLAIALPOVATITSPALPALASHWRLDATS
QMTALFFVGAPSGITVQASLRDVLQGPALMSYFTVIGAVIAMSPPVGLGQMLADIGH
LVQGLATCSVTYQASLRDVLQGPALMSYFTVIGAVIAMSPPVGLGQMLADIGH
PAVATVILASLAALVVPAMPTREPLAGTEPATLAFRRVADRPIOTRALLVA
VLNVLFSEYAGAPFVNGDLGLGFGWIGLAIATAGSLGALNNRRLPRTNSARVRL
GLALAGATVQOTLLAIVAGYAEGLYMLPALPIFIGGVAIPMLIGPALIAYDCCRR
AGALFGLAVYLLIGLIGASTLLPDRPWLAVWSLLGLVALLIARRLE"
2690..3469
/gene="PA1109"
2690..3469
/gene="PA1109"
/codon_start=1
/transl_table=11
/product="probable transcriptional regulator"
/protein_id="AAG04498.1"
/db_xref="GI:9947027"
/translation="MPRYPERVEDPDLARCPVGLVSDTGRSDMHCRRRAQLYM
AAGSVTLYFAERICQITLQOAWLPAGVPHRTVIGHFAFASLYFEPSAFVLPDQVO
VLEQPLMRLLIOMTGMADLOEGPHLTLRALPDELASPAHAASLIPMPDARL
KVAEALLAOPALMSTIEMAGREYAGSGRTIARLPFTQGTGFSRMTQARLFEARRRL
AEGRSYEVYHAHAYASDSAFITMYRRVCGQSGFRALRGLE"
complement(3474..4184)
/gene="PA1110"
complement(3474..4184)
/gene="PA1110"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04499.1"
/db_xref="GI:9947028"
/translation="MTDPIRLSKRLAELTACSRREAELEYEGGWVSVDSEVTEPOFK
VLDQRYVELPGARAETIEPATLILHRCAPMRHDDPGLIACRRMSDPSIRALKKH
FARQRPPLADEASGLVVEOOGHVLKRLVDGARLBOEYIVAVAGDIALACGLERKH
HGLAYOGRRLSPCKSWONESHRLFAKLVGLQGLRFMCEGRLVRSIRRLITALS
LALPLGEMRYGLIHERF"
4297..4920
/gene="PA1111"
4297..4920
/gene="PA1111"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04500.1"
/db_xref="GI:9947029"
/translation="MOSRLIDGLEMDLDEDSGGICLVNAGDKVHLROGDMQACGPGYC
IVMAMVQGLRRQAGLEIVDSRSRGKLMKALDQGLIRVGTGTVDLLELAEI
NTEHFAVLGEGRRLLVHRTREYLFAGFPVYIGFHRKQSDIRHMLATIGMSERATLL
DEAHLPRGOANNAVITSEAHSGRGYRIINAKGTWAVTLAEKVAL"
5013..6161
/gene="PA1112"
5013..6161

```

```

/gene="PA1112"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG04501.1"
/db_xref="GI:9947030"
/translation="MFGLSINAILLSAVALLPAYEAVDAGRHSDELGAVENTPVVGG
LAPVAVALEPDQGIIVTERPGSLAFVSPEQOLSAPEGVPRVPSAGGGLDVALS
PGFAEDRLVYLTYAEGEDGRAGTAAGRRLSADRKLGEFVIFRQPKLSSHHF
SRLVFDGKGLFTALGNNRSTODLDLQKLYRLYPDGGVPMDPNPVGADARPE
ISYGRNOGALANPWTGRIMTHHEGPRGDEINIPQAGVGMPIAHGNTYLLP
IPEARKAVGEPAPPHWEEKSPALSCAMPYAOGRPPAONHLLFGALATQELRLQL
DGRIVYHEERILCDLQVRLRDYRQGPDTYLLTLDQKDLRLGLAR"
7151..8917
/gene="PA1113"
7151..8917
/gene="PA1113"
/codon_start=1
/transl_table=11
/product="probable ATP-binding/permease fusion ABC
transporter"
/protein_id="AAG04502.1"
/db_xref="GI:9947031"
/translation="MPALISSRQRIALRLAMPENAPYRPMIGALVALLFTPAITLSL
GGGIRLVNPGGFLTASPELLNRYLLFTVYLGAVGTFFVFTVSWIGERVADIRK
RYFDHILIEHPFETENSSSEIOSRLADDTLLQTVIGSSLSMALRNMLVGVLLM
FTNKLKLSIVLALPLVLAIPILLFGRVRRLSRSQDRVADVSGYETIGQIKTVQ
AVNHQADRORFAGTVERAFDTRARRKIRQAMLLTVYLVVLAGVIMLVGGRDVA
GISISAGELAAEFVYSILVOMAEGLTSEVIELQRAAGAERIALELLRRSRTIRPAQ
RKLARVSGRIELQAVRFAYPTPRPMHIDISLAVEGTELLALVPGACKRSLFDL
ILRFDPDQOGRLLDQGPVLELDPADRLRRSRLRQALFPGVSEENIRGRPHASA
AVEFAARAHAHHEFTILPGQYKATHIEGGIGISGQOROLATIRALLVAPVLLD
EFTSALDAESEHLIDQALPGLMSGRTTLYIAHRLATVQSAURIAVTDGRLVAIGSHA
RLIAENPLIARIELQFAGGR"
complement(8920..9255)
/gene="PA1114"
complement(8920..9255)
/gene="PA1114"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04503.1"
/db_xref="GI:9947032"
/translation="MKHLAFLAGLILAPLAFAASPDCEKARADEKATCANALND
KDVMSYVRFRLRGIFAMGGRGAMQDRQOQMLRRRRGGGTDLCSRAIYORIDELDK
QYAGIDKPL"
complement(9359..11683)
/gene="PA1115"
complement(9359..11683)
/gene="PA1115"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04504.1"
/db_xref="GI:9947033"
/translation="MRPLSRWLSGCSLLAALFLVPLAGRYWLGMSPTFGYISDL
ALGSLVILLHRRSILFALPYMLMCFLIGNELVAAYARMEPSDLKTYLIDPOFVS
SSQGGIHLPLALGIGCATILCVILMRGRGRLPYVALPAAFLAAHVQIYLRP
SDADQOQVNLPHKILLASALSSGGLAYEDMLADRDPTPDVAGLQMLDLSGTLRG
GGRARNVLIYAMEGTPGATIAANRAAINSSYHEAPMRISGMAERAMTPDYVILSHQ
TIRGLIAMICGDSKLDSTGTPKAAVEILLNPARRECLPAOLRQHGSTERLQAGILRF
MAKDRVMPQMFQDTTIGDWFERNYSLEFPWGDMDCAFEAGATYVKOLROQKPYMLT
TILTVGHQPSYSAADYLARYPSAKOAAVYLDADDAIDAFLEDEKQGLVDTLVVLTLS
DESHGVQVRLASAMGFNVLAPEOALPPIKSGYGVHDLASVLDYGYRPAQIS
GSLFRDYASGRMWSYTNGLMRHHQKQDFVCDQOQCRRTRRSGFIADKEXEFTGR
FSGROALVYORAVLIDHSYQOQGOEOPATHTHERIRLKEPTHNDWMDNLLGAOYL
EMKRNSTTITLTKIRALQDRNCAVYLVTKTERENVPVLPVDPILRLGKPLDMSET
FDNEARKATVSHLAEGRGIIISDEHVVTANPVPVSRQVAAAREQOPLSTLTK
SLPANRETEARADAPRERYIIPQKALIIISH"
1889 a 3811 c 4174 g 1966 t

```

BASE COUNT
ORIGIN

ORGANISM cloning vector pBV1
artificial sequence

LPFLPDYKCKRLNSQVDLSNENYLOOLQLOEOMKKAADPFHAISTANGKARVGEAO
 PIKNPADHDIDVTGVSFADALAEAVGAANAPPEMSATPAEBAACLRPRADLEO
 HPALMLAVREAGKTINNATIEVREADPCRYVANEAEHTLPQAKVGAIVATISW
 NPLAIFTEGEVVSALAAGTVYAKAEOSTSLIAGASVLSMHEAGIPTSALDVLIGAG
 VCAALITNDARIGGVIFGSTEVARLLINKALRGDNPVLIETGOMAMIDSTALAE
 QVCAVLNAPFASGQCSALRLICVOEDVADRMIDMIKGAMDELVCKPIOLTVDVG
 PYIDAEQONLNLHINKMGVAKSYHEVKTADVDSKSTEVPRILFELNNINLEORE
 VEGPVLYVRYRADELNDYIDQINSKYALTGHVSRTEGTVRHRISRIEAGNVYVR
 NLVAVGVQOPRGHGLSGTGPKAGGSFYLQKTPRAGVWVAPTIISOQADBEALRL
 EALVHKLPNAEKKAAAAAGHARITRLRRETVLPTGTERNSISWARKRWIIG
 GSTVQAFALITELASGIOAVEPDSPLASTADIEGLLVNGKRPETAGISHVVALSP
 LDSARKQELAAHDGALIRLPSENGIDILQVEETISCSVNTTAAGNASLMAVAD"

gene
 complement(5628..7154)
 /gene="NMB0402"
 /note="similar to GB:L01146 GB:L01148 GB:L01136 GB:L01140
 GB:L01144 percent identity: 82.00; identified by sequence
 similarity: putative"
 /codon_start=1
 /transl_table=11
 /product="sodium/proline symporter"
 /protein_id="AAF40841.1"
 /db_xref="GI:7225624"
 /translation="NMPLYITFAIYLVAVLLIGLAAYFSTRNPDYILGKSGLPFVY
 AMSAGDSMGLMLGALYLSGLNEAMTAIGLVAGVFNMLIAGRLRVEHYAN
 NATLPDYFHRFGAGHLMKVVSALLILFFETTYCAGIVAGATLRPSLPEGMHYAN
 AMMIGAGATVYTLGFLVAVSMPTDLOASIMTALITTPVWYLGIGAGAKMSALIO
 SVAAGTCKEVSFAGTVIGITISAAAGLTFRGPHILAFKMAESKSLVSARIG
 MTWMAICLAAVAVGRTGIAFGAPDPVSSMSNHERIFALSTLFPNPIAGITIS
 ALIAVMSTLSQULLVCSSATTEDEYKGLERKNAQOSELVWVGRLMVAIVISILLA
 SDPSKVIIVSYAMAGFAFGPVIIVLSVLMKRITTAGALSGWAGALIVVWAEVY
 KKPAQAGESGLTMYEIVPGFIYCLIAVAVLSLKNKPSRETOERERKADADRIR

gene
 complement(7490..7582)
 /gene="NMB0403"
 /complement(7490..7582)
 /gene="NMB0403"
 /note="hypothetical protein; identified by Glimmer2;
 putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAF40842.1"
 /db_xref="GI:7225625"
 /translation="MPSENGRIIGKTYRASFAVAKLHKTPAA"

gene
 7637..7957
 /gene="NMB0404"
 /note="conserved hypothetical protein; identified by
 Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF40843.1"
 /db_xref="GI:7225626"
 /translation="MFCOLFEFEVSGKISFTIANSPAKDYENKIKAMLGAFNMDLY
 TRPEFDIOQVLTIRTKIALAEARLEKLEAQNPERAALEAAEAAEEVAELROOT
 EAGF"

gene
 7971..9467
 /gene="NMB0405"
 7971..9467
 /gene="NMB0405"
 /note="similar to GB:LA2023 SP:P45049 PID:1006428
 PID:1221236 PID:1205364 percent identity: 70.42;
 identified by sequence similarity: putative"
 /codon_start=1
 /transl_table=11
 /product="competence protein Comm"
 /protein_id="AAF40844.1"
 /db_xref="GI:7225627"
 /translation="MSLALVYSRLSGMNAFLVEVEAHLANGLPHEINIVGLPMEVKE

SRDVRRAIIIOGFEPPAKKTITVNLADLPKESGRFDLPDIAIGILASGOVAPKELE
 EYERAGELIASGLLRPRGALAMWQOMQKRAEVLREENAGAAWVRGTTVYGASL
 GEVAHLNGLEPLAOTECOVQPFEEGGQPDLDVGGQHTARLAEIAAGSHSLM
 MGPPGTGKSMISORLPGLILPILDELEVWALRSLLPNHOOLDNSRPRSRPHNS
 AAAMVGGSDPRPEITSLAHNGVFLFIDELPEFQDKVLEAREPLNENETIISAQOA
 VYRPFQLVAMNRCPCGICLGHVPRCPRTCPROGVAKRSKISGSLDRIDLTLEVSL
 SAALMOEGECSSASVLERVIARDKQYARQGVKNALSVSELDTSARIQKBOBAL
 GGLLEKISLSKRSFHRIMRVARTLADLADGEVGRSHVMAIGFRRL"

gene
 9567..10436
 /gene="NMB0406"
 9567..10436
 /gene="NMB0406"
 /note="conserved hypothetical protein; identified by
 Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAF40845.1"
 /db_xref="GI:7225628"
 /translation="MFNRKFSQSGKLGSEFPFLIATVYIAGILELYNGSGONAFKI
 PASSKOPATEILKPKMOPKEDIQPEPADNALSEPDATAEADQSDAKKADQPAD
 KADEVKAGEPEREPEPDGOAVRKALTEBERQTVREKQKDAETVTKQAVPSKET
 EKAKSEKKEKAAKEKVPKPTPOILNSGIERARSAAEVQKMTSDKAEATVYLO
 MGAYADROSAGORAKLALIGISKVYGVAGHKITLYRQSGMMSADAVKKMDELAK
 HEVASLIRSTESR"

gene
 10439..11083
 /gene="NMB0407"
 10439..11083
 /gene="NMB0407"
 /note="similar to SP:P95460 PID:181074 percent identity:
 56.84; identified by sequence similarity: putative"
 /codon_start=1
 /transl_table=11
 /product="thiol:disulfide interchange protein Dsba"
 /protein_id="AAF40846.1"
 /db_xref="GI:7225629"
 /translation="MKIKHLPLLSAVLSAQAVALTEGEDVYLIDKPIPOSGKTE
 VLEFGYFCVCHNRPRLILKLKALPSATYLTETVYQNPQREMLGARAANVLSGL
 KYQANRAVRAVYEQKTRLENSRVAGKMAVLSQGFQCKLMRAVYDSEPAALAKMKQK
 LTEQYRIDSTPRTYIVGSKYRVIFNNGEDGVITIKELVAKVREERKQTPAVQK"

gene
 11116..11937
 /gene="NMB0408"
 11116..11937
 /gene="NMB0408"
 /note="similar to PID:882579 GB:U00096 PID:1789437 percent
 identity: 67.54; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="bacitracin resistance protein"
 /protein_id="AAF40847.1"
 /db_xref="GI:7225630"
 /translation="MDFLIVLAKMLMGIVEGTFEPLPISSTGHLVFGNLIQFHSNHK
 VPEIATOLGAVLAIVEYRORFSNVJHGLGKRRKARFVLAIAFIPAAVNGLEK
 QIKVEYFNPLSVAVMLVIGFETLWEEKRSRAEPRIADVADLRPIDMLIGVAYFA
 LVGTSRSRSTINGMLMGIERTKTAEPSEFPLAVPRMAVAATVADLKHKRFTLNDVG

Query Match 71.28; Score 17.8; DB 1; Length 12107;
 Best Local Similarity 90.5%; Pred. No. 7.1e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 catcgacgacatctgacct 22
 Db 10601 CATTGCCATCATTTGCATCCT 10621

RESULT 38
 AC018306 14645 bp DNA HTG 09-DEC-1999
 LOCUS AC018306/c
 DEFINITION Dirosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
 pieces.
 ACCESSION AC018306
 VERSION AC018306.1 GI:6552885

KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 14645)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10214280 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. 14645
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 4097 a 3678 c 3403 g 3467 t
ORIGIN
Query Match 71.2%; Score 17.8; DB 65; Length 14645;
Best Local Similarity 90.5%; Pred. No. 7e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 gcatgcacagcattcgatcc 21
|||||
Db 10697 GCATGCCAGCATTCGCATCC 10677
AP003375 106462 bp DNA HTG 27-MAR-2001
LOCUS Oryza sativa chromosome 1 clone OJ1414_E05, *** SEQUENCING IN
DEFINITION PROGRESS ***, in ordered pieces.
ACCESSION AP003375
VERSION AP003375.1 GI:13365592
KEYWORDS HTG: HTGS_PHASE2.
SOURCE Oryza sativa (cultivar: Nipponbare) DNA, clone: OJ1414_E05.
ORGANISM Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
Oryza.
REFERENCE 1 (sites)
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, BAC
clone: OJ1414_E05
JOURNAL Published only in Database (2001) In press
AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2001) Takuji Sasaki, National Institute of
Agricultural Resources, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(e-mail: tsasaki@ab.affrc.go.jp, URL: http://rgrp.dna.affrc.go.jp/
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
COMMENT The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces between them
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers

source
1. 106462
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="OJ1414_E05"
BASE COUNT 28948 a 24036 c 23776 g 29652 t 50 others
ORIGIN
Query Match 71.2%; Score 17.8; DB 83; Length 106462;
Best Local Similarity 90.5%; Pred. No. 6.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 gcatgcacagcattcgatcc 21
|||||
Db 14284 GCATGCCAGCATTCGCATCC 14264
RESULT 40
TBCHRIA_03
WPCOMMENT
Sequence split into 10 fragments LOCUS TBCHRIA Accession AL359782
Fragment Name Begin End
TBCHRIA_00 1 110000
TBCHRIA_01 100001 210000
TBCHRIA_02 200001 310000
TBCHRIA_03 300001 410000
TBCHRIA_04 400001 510000
TBCHRIA_05 500001 610000
TBCHRIA_06 600001 710000
TBCHRIA_07 700001 810000
TBCHRIA_08 800001 910000
TBCHRIA_09 900001 941631
Continuation (4 of 10) of TBCHRIA from base 300001 (AL359782 Trypanosoma brucei chro
Query Match 71.2%; Score 17.8; DB 84; Length 110000;
Best Local Similarity 90.5%; Pred. No. 6.5e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 gcatgcacagcattcgatcc 21
|||||
Db 41181 GCATGCCAGCATTCGCATCC 41201
RESULT 41
AC007809
LOCUS AC007809 183439 bp DNA INV 24-FEB-2001
DEFINITION Drosophila melanogaster, chromosome 3R, region 88C-88C, BAC clone
BACR45M03, complete sequence.
ACCESSION AC007809
VERSION AC007809.8 GI:13122709
KEYWORDS HTG:
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 183439)
AUTHORS Gelniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,
Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
Rogers, Y., Ab, H., Baldwin, D., Bantz, J., Beeson, K.Y., Busam, D.A.,
Carlson, J.W., Center, A., Chapple, M., Davenport, L.B., Dietz, S.M.,
Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,
Fieriera, S., Frisoe, E., Galle, R.F., Garg, N.S., George, R.A.,
Gonzalez, M., Houch, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
Ibenwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A.,
McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nuno, J.,
Paclet, J., Paragov, V., Park, S., Patel, S., Pfeiffer, B.,
Phonananavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
Stapleton, M., Strong, R., Swirekas, R., Tector, C., Williams, S.M.,
Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

[illegible][illegible]

misc_feature	complement(4262..4271)
	/note="Core DNA uptake sequence: gccgctcga"
gene	/label=DUS complement(4298..5497) /gene="ackA2" complement(4298..5497) /gene="ackA2" EC_number=2.7.2.1"
CDS	/note="NMA2050; ackA2, acetate kinase, len: 399aa; similar to many eg. SW:P15046 (ACKA_ECOLI) ackA, acetate kinase from Escherichia coli (400 aa) fasta scores: E(): 0.49, 55 identity in 400 aa overlap. Also similar to NMA1718 from Neisseria meningitidis fasta scores: E(): 0.63, 38 identity in 398 aa overlap. Consensus Pfam match to entry PF00871 Acetate_kinase, Acetokinas family; Prosite match to PS01075 Acetate and butyrate kinases family signature 1 and Prosite match to PS01076 Acetate and butyrate kinases family signature 2." /codon_start=1 /transl_table=11 /product="acetate kinase" /protein_id="CAB85268.1" /db_xref="GI:7380678" /db_xref="SPTRMBL:Q9JY07" translation="MSDQLILVNLGSSSLKGAVIDRKSGSVLSICGERLTTPPAVTESDGNGRROYPLSGRNAGAVAGMLNELKEHLIDIAQWGRHRIANGKRSYSVIDAWDELNAFCIPALPHANPANSIGLIAADHEHPGLDNVAMDTSTHOTPERVTVAYPRRLKRKYAFRRIRGHGTSMRYVAPEACTITCKPLPDIMTIIAHLGNSITAIKRGKSVDSMGFTPIELGVWGTRCGIDDPGVSYLSHAIDLVAQVDENMKSGSLGISLRLSDCRTLEIAADEGEGCARLALEMVTRYAKTIASNAVGCQVDALVFYGISENRINAKTYVSYLDPLGLHDITKANNEKRYGNISITPSDSPAVLVPTNELMIACDI AELVGIL"
misc_feature	complement(4328..5485)
Query Match	71.2%; Score 17.8; DB 3; Length 195767;
Best Local Similarity	90.5%; Pred. No. 6.4e+02;
Matches 19; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
Oy	2 catcgccagcatlccatctt 22 Db 38896 CATGCCATCATTTGGATCT 38876
RESULT 43	
AEO03705	AEO03705 232744 bp DNA INV 05-OCT-2000
LOCUS	Drosophila melanogaster genomic scaffold 142000013386035 section 30
DEFINITION	OF 105, complete sequence.
ACCESSION	AEO03705 AF002708
VERSION	AEO03705.1 GI:7299886
KEYWORDS	fruit fly.
SOURCE	Drosophila melanogaster
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 232744) Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanalides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Brandon,R.C., Rogers,Y.H., Blazer,R.G., Champe,M., Pfeiffer,B.D., Wen,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabler,Miklos,G.L., Abail,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A., Bendale,J., Bonckar,W., Bork,P., Bhandari,D., Boldakov,S., Borokova,D., Botchan,M.R., Bouck,J., Brockstein,P., Brotilier,P., Burris,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dunn,P., Durbin,K.I., Evans,R.L., Favelle,R., Fertig,C., Flaxman,C., Dunn,P., Durbin,K.I., Evans,R.L., Favelle,R., Fertig,C., Flaxman,C.,
REFERENCE	
AUTHORS	

	FEATURES	
	Source	Location/Qualifiers
	1. 332744	
	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
	/chromosome="3R"	
	complement(join(29274..29439,29524..30038,30107..>30980))	
mRNA	/gene="CG8464"	
	/product="CT24781"	
	/db_xref="FLYBASE:FBan0008464"	
	/db_xref="FLYBASE:FBgn0038233"	
	complement(<29274..>30980)	
	/gene="CG8464"	
	/map="88C4-88C6"	
	/db_xref="FLYBASE:FBan0008464"	
	/db_xref="FLYBASE:FBgn0038233"	
	complement(join(29274..29439,29524..30038,30107..30694))	
CDS	/gene="CG8464"	
	/note="CG8464 gene product"	
	/codon_start=1	
	/db_xref="FLYBASE:FBan0008464"	
	/db_xref="FLYBASE:FBgn0038233"	
	/protein_id="AAF55062.1"	
	/db_xref="GI:7299887"	
	/translation="MALRGSRLVLEIFKRCIASPVLIHSANRRSSOLAIKEGPN GNGQYQNGEOKGWRLEVEFVSLGAVNSAIIQREDLPTIASKGRRRD FNIAIYVAGCDSVYLEIKDTRHDFSGQPIIASNGSGFIIEQNLIILTHAVI NKHHTAVQRLSGRTFPAIEDVDQSLALRLIOVNNLSYMRIGKSGTIRSGWV ALGSPALINWTYAGVISTORASOELGRNDIINLOTDAITFGNSGSGPLVINDGE AICGNSMKYTAGISFALPIDYVKVLEPRAEKRRKGSAYKTCGYPKRRMGITMLTTP DILFELKRSQMPNSLTHGVLIKVTGSPAHSGLDGDIIVTHINKKEIKNSDYI DALDSSKTLIDVILKGVKQMHVTTIPEDP"	
mRNA	join(<31212..31346,31407..31691,31752..>31922)	
	/gene="CG3351"	
	/product="CT11255"	
	/db_xref="FLYBASE:FBan0003351"	
	/db_xref="FLYBASE:FBgn0038234"	
	<31212..>31922	
gene	/gene="CG3351"	
	/map="88C4-88C6"	
	/db_xref="FLYBASE:FBan0003351"	
	/db_xref="FLYBASE:FBgn0038234"	
	join(31212..31346,31407..31691,31752..31922)	
CDS	/gene="CG3351"	
	Flatschmann W., Foster C., Giorliani A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibbewam C., Jafari M., Kalush F., Karpén G.H., Ke Z., Kernison S., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Krautitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li D., Li Z., Liang Y., Lin X., Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M., Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wasserman D.A., Weisenbach G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Zhong R.A., Myers E.W., Rubin G.M. and Venter C.J. The genome sequence of Drosophila melanogaster Science 287 (5461), 2185-2195 (2000)	
TITLE	2 (bases 1 to 232744)	
MEDLINE	2185-2195 (2000)	
REFERENCE	Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M. and Venter C.J. Direct Submission Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA	
AUTHORS	Rockville, MD, USA	
JOURNAL	Location/Qualifiers	
FEATURES	1. 332744	
Source	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
	/chromosome="3R"	
	complement(join(29274..29439,29524..30038,30107..>30980))	
mRNA	/gene="CG8464"	
	/product="CT24781"	
	/db_xref="FLYBASE:FBan0008464"	
	/db_xref="FLYBASE:FBgn0038233"	
	complement(<29274..>30980)	
	/gene="CG8464"	
	/map="88C4-88C6"	
	/db_xref="FLYBASE:FBan0008464"	
	/db_xref="FLYBASE:FBgn0038233"	
	complement(join(29274..29439,29524..30038,30107..30694))	
CDS	/gene="CG8464"	
	/note="CG8464 gene product"	
	/codon_start=1	
	/db_xref="FLYBASE:FBan0008464"	
	/db_xref="FLYBASE:FBgn0038233"	
	/protein_id="AAF55062.1"	
	/db_xref="GI:7299887"	
	/translation="MALRGSRLVLEIFKRCIASPVLIHSANRRSSOLAIKEGPN GNGQYQNGEOKGWRLEVEFVSLGAVNSAIIQREDLPTIASKGRRRD FNIAIYVAGCDSVYLEIKDTRHDFSGQPIIASNGSGFIIEQNLIILTHAVI NKHHTAVQRLSGRTFPAIEDVDQSLALRLIOVNNLSYMRIGKSGTIRSGWV ALGSPALINWTYAGVISTORASOELGRNDIINLOTDAITFGNSGSGPLVINDGE AICGNSMKYTAGISFALPIDYVKVLEPRAEKRRKGSAYKTCGYPKRRMGITMLTTP DILFELKRSQMPNSLTHGVLIKVTGSPAHSGLDGDIIVTHINKKEIKNSDYI DALDSSKTLIDVILKGVKQMHVTTIPEDP"	
mRNA	join(<31212..31346,31407..31691,31752..>31922)	
	/gene="CG3351"	
	/product="CT11255"	
	/db_xref="FLYBASE:FBan0003351"	
	/db_xref="FLYBASE:FBgn0038234"	
	<31212..>31922	
gene	/gene="CG3351"	
	/map="88C4-88C6"	
	/db_xref="FLYBASE:FBan0003351"	
	/db_xref="FLYBASE:FBgn0038234"	
	join(31212..31346,31407..31691,31752..31922)	
CDS	/gene="CG3351"	
	Flatschmann W., Foster C., Giorliani A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibbewam C., Jafari M., Kalush F., Karpén G.H., Ke Z., Kernison S., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Krautitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li D., Li Z., Liang Y., Lin X., Liu X., Maltel B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Pacleb J.M., Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wasserman D.A., Weisenbach G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S	

LCGMMTACDVSAIAKPWEYOHKAKLVADDEFDQGLEKLOJNTOPVMMRERKD
 ELPRKQVGFIDVLCPLRYVLCDFPWTJYEGTLENNMMDLAKVMGJLWIDH
 DTIDKPEEFACACADEIKDIETVTLLNNOGSHSHTPEHRRSSRLSMKKT
 GALGKAVRSKSLKLYNSMDGSKPKTSLKLEISVSDMDKSPITSPOQASGSMGR
 MSASSTSSAGAGVNDKSKRSKLCALD.
 join(40359, .40547,40626, .40826,40891, .41076,41138, .41210,
 .41284, .41410,41461, .42034,42102, .>42262)
 /gene="Cyp313a1"
 /product="CT11271"
 /db_xref="FLYBASE:FBann003360"
 /db_xref="FLYBASE:FBgn0038236"
 <40359, >42262
 /gene="Cyp313a1"
 /note="CG3360"
 /map="88C5-88C6"
 /db_xref="FLYBASE:FBann003360"
 /db_xref="FLYBASE:FBgn0038236"
 join(40391, .40547,40626, .40826,40891, .41076,41138, .41210,
 .41284, .41410,41461, .42034,42102, .42262)
 /gene="Cyp313a1 gene product"
 /note="Cyp313a1"
 /db_xref="FLYBASE:FBann003360"
 /db_xref="FLYBASE:FBgn0038236"
 /protein_id="AA55065.2"
 /db_xref="GI:10726350"

Query Match 71.2%; Score 17.8; DB 5; Length 232744;
 Best Local Similarity 90.5%; Pred. No. 6.3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 106436 CCTGCCAGCATTTGCATCAT 106456
 1 |||||||||||||||

RESULT 44
 AC017740/c
 LOCUS
 DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
 pieces.
 AC017740 237119 bp DNA HTG 10-DEC-1999
 AC017740.1 GI:6554259
 HTG: HTGS_PHASE2.
 SOURCE fruit fly
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 237119)
 Adams, M. and Venter, J.C.
 REFERENCE
 AUTHORS Direct Submission
 TITLE Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
 JOURNAL Rockville, MD, USA
 COMMENT This sequence was identified as CDM:10211941 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1..237119
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"

BASE COUNT 67689 a 49909 c 50172 g 69349 t
 ORIGIN

Query Match 71.2%; Score 17.8; DB 65; Length 237119;
 Best Local Similarity 90.5%; Pred. No. 6.3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 catgccagcattgcattcct 22

Db 103683 CCTGCCAGCATTTGCATCAT 103663
 1 |||||||||||||||

RESULT 45
 AC005557
 LOCUS
 DEFINITION Drosophila melanogaster, chromosome 3L, region 62A10-62B5, PI
 clones DS02777, DS03222, DS02345, and DS04808, complete sequence.
 AC005557 AC004346 AC004344 AC004354 AC004332
 AC005557.1 GI:3492862
 HTG.
 SOURCE fruit fly
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 257867)
 Celiker, S.E., Abayani, A., Arcalata, T.T., Baxter, E., Blazej, R.G.,
 Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomolan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Punch, E., Sequeira, A., Sethi, H., Snir, E.,
 Svitskas, R.R., Twomey, B., Wan, K.H., Weinburg, T., Zhang, R.,
 Zieran, L.L. and Rubin, G.M.
 REFERENCE
 AUTHORS Sequencing of Drosophila chromosome 3L, region 62A10-62B5
 JOURNAL Unpublished (1998)
 TITLES
 JOURNAL 2 (bases 1 to 257867)
 REFERENCE
 AUTHORS Celiker, S.E., George, R.A., Galle, R.F., Hoskins, R.A.,
 Svitskas, R.R., Harris, N.L., Abayani, A., Arcalata, T.T., Baxter, E.,
 Blazej, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E.,
 Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, S.H., Lee, B., Lomolan, M.A., Mak, J., Mazda, P., Mok, M.S.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
 Pfeiffer, B., Punch, E., Snir, E., Twomey, B., Wan, K.H., Whitehead, K.R.,
 Yee, A., Zhang, R., Zieran, L.L. and Rubin, G.M.
 TITLES
 JOURNAL Direct Submission
 JOURNAL Submitted (29-AUG-1998) Berkeley Drosophila Genome Project, MS
 64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
 Berkeley, CA 94720, US
 COMMENT Sequence submitted by:
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu.
 PI library locations: 29-89, 34-54, 25-41.

FEATURES
 source
 1..257867
 /organism="Drosophila melanogaster"
 /strain="Y2: cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="3L"
 /map="62A10-62B5"
 /clone_id="PI library, partial Sau3a in pAd10SACBIT"
 /clone="Pis DS02777 (D245), DS03222 (D243), DS02345
 (D257), and DS04808 (D228)"
 /note="DS02777 (D245) extends from bp 1 to bp 78,221.
 DS03222 (D243) is a bridge PI extending from bp 78,210 to
 PI end at bp 142,019. DS02345 (D257) is a bridge PI
 extending from bp 141,146 to bp 167,538. DS04808 (D228)
 extends from bp 170,514 to bp 257,867."

BASE COUNT 66126 a 61150 c 61220 g 69371 t
 ORIGIN

Query Match 71.2%; Score 17.8; DB 4; Length 257867;
 Best Local Similarity 90.5%; Pred. No. 6.3e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

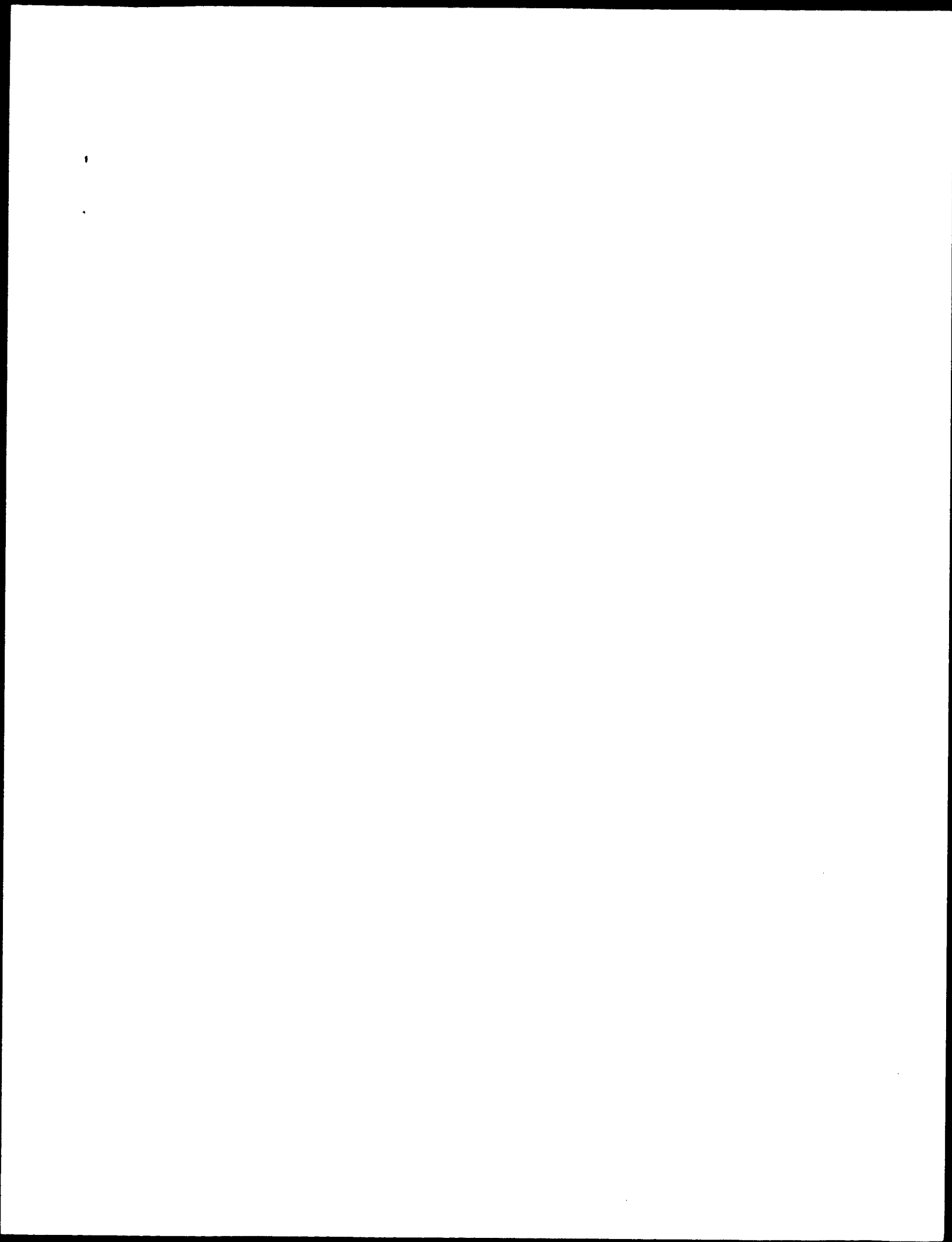
Wed Oct 10 07:44:47 2001

us-09-396-196f-7.std.rge

Page 39

Qy 1 gcacgcagcattcgatcc 21
|||||
Db 49770 GCATCGCCGATCGGATCC 49790

Search completed: October 9, 2001, 12:13:27
Job time: 3922 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:38:55 ; Search time 470.56 Seconds
(without alignments)
33.359 Million cell updates/sec

Title: US-09-396-196F-7

Perfect score: 25

Sequence: 1 gcacgcgcagcattcgatccctcgt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 73101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0601.*

1:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1980.DAT.*
2:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1981.DAT.*
3:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1982.DAT.*
4:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1983.DAT.*
5:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1984.DAT.*
6:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1985.DAT.*
7:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1986.DAT.*
8:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1987.DAT.*
9:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1988.DAT.*
10:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1989.DAT.*
11:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1990.DAT.*
12:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1991.DAT.*
13:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1992.DAT.*
14:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1993.DAT.*
15:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1994.DAT.*
16:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1995.DAT.*
17:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1996.DAT.*
18:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1997.DAT.*
19:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1998.DAT.*
20:	/SIDSI/gcgcdata/geneseq/geneseqn/NA1999.DAT.*
21:	/SIDSI/gcgcdata/geneseq/geneseqn/NA2000.DAT.*
22:	/SIDSI/gcgcdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	25	100.0	1041	20	AAAX01303
2	25	100.0	1084	10	AAAN91329
3	25	100.0	1121	7	AAAN60496
4	25	100.0	5872	15	AAO62386
5	18.2	72.8	3451	18	AAV02308
6	17.8	71.2	645	21	AAZ54339
7	17.8	71.2	645	21	AAZ54340
8	17.8	71.2	645	21	AAZ54341
9	17.8	71.2	69936	21	AAAB1479
10	17.8	71.2	349980	21	AAAF21607
11	17.8	71.2	1437668	21	AAAB1490

Result ID	Score	Match	Query Length	DB ID	Description	ALIGNMENTS
1	25	100.0	1041	20	AAAX01303	AAAX01303 standard; DNA; 1041 BP.
2	25	100.0	1084	10	AAAN91329	
3	25	100.0	1121	7	AAAN60496	
4	25	100.0	5872	15	AAO62386	
5	18.2	72.8	3451	18	AAV02308	
6	17.8	71.2	645	21	AAZ54339	
7	17.8	71.2	645	21	AAZ54340	
8	17.8	71.2	645	21	AAZ54341	
9	17.8	71.2	69936	21	AAAB1479	
10	17.8	71.2	349980	21	AAAF21607	
11	17.8	71.2	1437668	21	AAAB1490	

XX Example 2; Column 37-40; 34pp; English.

PS This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC be used in the transgenic plant of the invention. The transgenic plant,
CC plant cell or plant tissue is transformed with a chimeric gene encoding
CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC produces more biotin than a non-transgenic plant, cell or tissue. The
CC plant is used as an improved dietary source of biotin (vitamin H) for
CC humans or animals.

XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.009;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatcgccagcattcgatcctcgt 25
Db 90 gcatcgccagcattcgatcctcgt 114

RESULT 2

AA062386 standard; DNA; 1084 BP.

AA062386;

15-FEB-1990 (first entry)

E. coli Bio B gene.

E. coli Bio B gene; biotin.

Escherichia coli.

Key Location/Qualifiers

FT CDS 24..1064

FT /*tag=a

GB2216530-A.

11-OCT-1989.

17-MAR-1989; 89GB-0006210.

22-MAR-1988; 88GB-0006804.

17-MAR-1989; 89GB-0006210.

(UKAG-) UK MIN. AGRIC. FTSH.

Pearson BM, McKee RA;

WPI; 1989-295085/41. P-PSDB P91392

Plasmid contg. gene(s) for expression of biotin synthetase enzymes

derived from E.coli and capable of replication and expression in other

microorganisms, esp. yeast.

Table 3; page 33-4; 52pp; English.

The gene can be used in a plasmid for expression of enzymes of the biotin

synthetic pathway. Pref. control sequences for expression in S.cerevisiae

are plasmids pMA91, pMA36C, pKV49 and pCK495, and plasmid pCK965 for

Lactobacillus. Insertion of bio B improves biotin yields in

microorganisms which export biotin, or enables growth in media contg.

little or no biotin of organisms unable to synthesise biotin for their

Query Match 100.0%; Score 25; DB 10; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatcgccagcattcgatcctcgt 25
Db 113 gcatcgccagcattcgatcctcgt 137

RESULT 3

AA060496 standard; DNA; 1121 BP.

AA060496;

17-OCT-1991 (first entry)

Sequence encoding biotin synthesising enzyme.

Biotin synthetic enzyme; E.coli; deschiobiotin; ds.

Key Location/Qualifiers

FT CDS 42..1082

FT /*tag=a

JP61149091-A.

07-JUL-1986.

24-DEC-1984; 84JP-0272605.

24-DEC-1984; 84JP-0272605.

(NIPS) NIPPON SODA KK.

WPI; 1986-216622/33.

P-PSDB; AAP60536.

Double stranded DNA encoding biotin synthesising enzyme -

comprises transformed mutant E.coli strain contg. cyclic doubled

stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

Disclosure; Page 534; 23pp; Japanese.

The sequence may be expressed by a transformed E.coli host, cultured

in a medium containing deschiobiotin.

Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;

Best Local Similarity 100.0%; Pred. No. 0.0091;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatcgccagcattcgatcctcgt 25

Db 131 gcatcgccagcattcgatcctcgt 155

RESULT 4

AA062386 standard; DNA; 5872 BP.

AA062386;

16-NOV-1994 (first entry)

Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

Biotin; expression; enterobacteria; vitamin H; synthesis;

plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;

promoter plac; biotin synthase; KAPA synthase;

8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;


```

FH Key Location/Qualifiers
FT CDS 340..3207
FT /tag= a
FT /product= "cell membrane proton-ATPase"
XX
XX JP09252786-A.
XX
XX 30-SEP-1997.
XX
XX 25-MAR-1996; 96JP-0124124.
XX
PR 25-MAR-1996; 96JP-0124124.
XX
XX (ORLY ) ORIENTAL YEAST CO LTD.
XX
XX MPI: 1997-530157/49.
XX DR P-PDB; AAM31363.
XX
XX Proton-ATPase and cDNA coding sequence - useful for producing
PT acid-resistant transgenic plants
XX
PS Claim 1; Page 11-13; 16pp; Japanese.
CC
CC The present sequence encodes a cell membrane proton-ATPase.
CC The gene is useful for producing acid-resistant transformed
CC plants. It also allows large scale production of the enzyme.
XX
SQ Sequence 3451 BP; 818 A; 860 C; 960 G; 813 T; 0 other;

Query Match 72.8%; Score 18.2; DB 18; Length 3451;
Best Local Similarity 87.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 gcacgcacgacattcgaccctc 23
    ||| ||||| | | |||||
Db 812 GCATTGCCAGCATCGCTTCTC 790

RESULT 6
AAZ54339
ID AAZ54339 standard; DNA; 645 BP.
XX
XX AAZ54339;
AC
XX
XX 21-MAR-2000 (first entry)
DT
XX
DE Neisseria gonorrhoeae ORF 767 partial DNA sequence SEQ ID NO:2627.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy; ds.
XX
OS Neisseria gonorrhoeae.
XX
PN W09957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
```

PI	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI	Petersen J, Pizsa M, Rappoli R, Ratti G, Scalato E, Scarselli M;
PI	Tettelein H, Venter JC;
XX	
DR	WPI; 2000-062150/05.
XX	
XX	P-PsDB; AAY75577.
PT	Novel Neisserial polypeptides predicted to be useful antigens for
PT	vaccines and diagnostics _
XX	
PS	Claim 7; Page 1246; 1453pp; English.
XX	
CC	AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC	represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC	and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
CC	PCR primers used in the exemplification of the present invention. The
CC	polypeptides, the polynucleotides, antibodies and compositions of
CC	the invention can be used as vaccines, as diagnostic reagents, and as
CC	immunogenic compositions. The polypeptides can be used in the
CC	manufacture of medicaments for treating or preventing infection due to
CC	Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC	presence of Neisseria bacteria, or to raise antibodies. They may also
CC	be used to screen for agonists or antagonists, which may themselves
CC	have use as antibacterial agents. The polynucleotides of the invention
CC	may also be used in gene therapy protocols.
XX	
SQ	Sequence 645 BP; 170 A; 138 C; 171 G; 166 T; 0 other;
Query Match	71.2%; Score 17.8; DB 21; Length 645;
Best Local Similarity	90.5%; Pred No.22;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0	
OY	2 catgccacgattgcgatcct 22
Db	163 catggccatcatcgatcct 183
RESULT 7	
AAZ54340	ID AAZ54340 standard; DNA; 645 BP.
XX	
AC	AAZ54340;
XX	
DT	21-MAR-2000 (first entry)
XX	
DE	Neisseria meningitidis ORF 767 partial DNA sequence SEQ ID NO:2629.
XX	
KM	Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW	antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW	antibacterial; gene therapy; ds.
XX	
OS	Neisseria meningitidis.
XX	
PN	WO957280-A2.
XX	
PD	11-NOV-1999.
XX	
PF	30-APR-1999; 99WO-US09346.
XX	
PR	01-MAY-1998; 98US-0083758.
PR	31-JUL-1998; 98US-0094869.
PR	02-SEP-1998; 98US-0098994.
PR	02-SEP-1998; 98US-0099062.
PR	09-OCT-1998; 98US-0103749.
PR	09-OCT-1998; 98US-0103794.
PR	09-OCT-1998; 98US-0103796.
PR	25-FEB-1999; 99US-0121528.
XX	
PA	(CHIR) CHIRON CORP.
PA	(GENO-) INST GENOMIC RES.
XX	
PI	Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M,

PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 DR WPI: 2000-062150/05.
 DR P-PSDB: AAY75578.
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 PS Claim 7: Page 1247; 1453pp; English.
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoea polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 SQ Sequence 645 BP; 168 A; 147 C; 170 G; 160 T; 0 other;

Query Match 71.2%; Score 17.8; DB 21; Length 645;
 Best Local Similarity 90.5%; Pred. No. 22;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 catgcgcagcattcgatcct 22
 ||| |||| ||||| ||||| |||||
 Db 163 catgcgcattcgatcct 183

RESULT 8
 ID AA254341 standard; DNA: 645 BP.
 AC AA254341;
 XX
 DT 21-MAR-2000 (first entry)
 XX
 DE Neisseria meningitidis ORF 767 partial DNA sequence SEQ ID NO:2631.
 XX
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KM antibacterial; gene therapy; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PE 30-APR-1999; 99WO-US09346.
 XX
 PF 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GBNO-) INST GENOMITC RES.
 XX
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR P-PSDB: AAY75579.
 XX
 PT Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 PS Claim 7: Page 1247-1248; 1453pp; English.
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoea polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 SQ Sequence 645 BP; 170 A; 146 C; 170 G; 159 T; 0 other;

Query Match 71.2%; Score 17.8; DB 21; Length 645;
 Best Local Similarity 90.5%; Pred. No. 22;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 catgcgcagcattcgatcct 22
 ||| |||| ||||| ||||| |||||
 Db 163 catgcgcattcgatcct 183

RESULT 9
 ID AAA81479/C
 XX AAA81479 standard; DNA: 69936 BP.
 AC AAA81479;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE N. meningitidis partial DNA sequence gnm_27 SEQ ID NO:27.
 XX
 KM Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
 KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KM Meningococcus B; MenB; ds.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO200022430-A2.
 XX
 PD 20-APR-2000.
 XX
 PE 08-OCT-1999; 99WO-US23573.
 XX
 PF 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Fraser CM, Hickey E, Peterson J, Tettelin H, Venter JC;
 PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
 PI Rappuoli R, Pizza M;
 DR WPI: 2000-318079/27.
 XX
 PT Isolated nucleotide sequences of Neisseria meningitidis which can be
 PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -

PS Claim 7; Page 547-567; 1760pp; English.

CC The present invention describes methods of obtaining immunogenic
 CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed *Neisseria meningitidis* genomic DNA
 CC sequences. AAA81560 to AAA81303 and AAB25620 to AAB25663 represent
 CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to
 CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC *Neisseria* bacteria. For example, some of the identified proteins could
 CC be components of vaccines against *Meningococcus B*; against all serotypes;
 CC and/or against all pathogenic *Neisseriae*. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

SO Sequence 69936 BP; 16512 A; 19259 G; 17399 G; 16763 T; 3 other;

Query Match 71.2%; Score 17.8; DB 21; Length 69936;
 Best Local Similarity 90.5%; Pred. No. 40;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 catgccagcattcgatcct 22
 ||| |||| ||||| ||||| |||||
 Db 66524 CATGCCATCATTCGATCCT 66504

RESULT 10

AAF21607
 ID AAF21607 standard; DNA; 349980 BP.

AC AAF21607;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE *Neisseria meningitidis B* nucleotide sequence SEQ ID NO:108.
 XX
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; immunogenic; vaccine;
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 XX ds.
 XX *Neisseria meningitidis*.
 OS
 PN WO20006791-A1.
 XX
 PD 09-NOV-2000.
 XX
 PF 08-MAR-2000; 2000WO-US05928.
 XX
 PR 30-APR-1999; 99US-0132068.
 PR 08-OCT-1999; 99WO-US23573.
 PR 28-FEB-2000; 2000GB-0004695.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Pizze M, Hickey E, Peterson J, Tettelin H, Venter JC, Masiagnani V;
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V, Rappuoli R;
 PI Frazer CM, Grandi G;
 XX

DR WPI; 2000-647603/62.

XX *Neisseria meningitidis B* full length genome sequence and open reading
 PT frames are used to detect, treat and prevent *Neisseria* infections -
 XX

PS Claim 7; Appendix A; 692pp; English.

CC The present invention describes the full length genome of
 CC *Neisseria meningitidis B* (NMB). The sequences in AAF21544 and AAF21607
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the
 CC sequence was too long to go in a record on its own. It was split into 8
 CC sequences which overlap each other at the beginning and end of each
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
 CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
 CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the
 CC *Neisseria* proteins given in AAB58550 to AAB58593, and AAF21589 to
 CC AAF21606 represent PCR primers which are used in the exemplification of
 CC the present invention. The NMB genome and fragments from it have
 CC antibacterial activity, and can be used in vaccines and gene therapy.
 CC *Neisseria* nucleic acids, proteins and/or antibodies which binds to the
 CC proteins can be used in compositions for treating or preventing infection
 CC due to *Neisseria* bacteria or as a diagnostic reagent for detecting the
 CC presence of *Neisseria* bacteria or of antibodies raised to *Neisseria*
 CC bacteria. Computers, computer memory, computer storage medium or computer
 CC databases can be used in a search to identify open reading frames (ORFs)
 CC or coding sequences within the NMB genome. The DNA sequences provide
 CC further opportunities to find antigenic or immunogenic proteins which are
 CC more effective in vaccines than the outer membrane proteins currently
 CC used.

SO Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

Query Match 71.2%; Score 17.8; DB 21; Length 349980;
 Best Local Similarity 90.5%; Pred. No. 48;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 catgccagcattcgatcct 22
 ||| |||| ||||| ||||| |||||
 Db 114663 catgccagcattcgatcct 114683

RESULT 11

AAA81490
 ID AAA81490 standard; DNA; 1437668 BP.

AC AAA81490;
 XX
 DT 04-DEC-2000 (first entry)
 XX
 DE *N. meningitidis B* full length genome DNA sequence SEQ ID NO:1068.
 XX
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
 KW *Meningococcus B*; MenB; ds.
 XX
 XX *Neisseria meningitidis*.
 OS
 PN WO200022430-A2.
 XX
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999; 99WO-US23573.
 XX
 PR 09-OCT-1998; 98US-0103794.
 PR 30-APR-1999; 99US-0132068.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC,
 PI Masiagnani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;
 PI Rappuoli R, Pizze M;
 XX

DR WPI; 2000-318079/27.

XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be

PT used in the diagnosis and treatment of *N. meningitidis* infection and

XX other *Neisseria* infections, for example, *N. gonorrhoea* -

PS Claim 7; Page 866-1272; 1760pp; English.

XX The present invention describes methods of obtaining immunogenic

CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414

CC represent specifically claimed *Neisseria meningitidis* genomic DNA

CC sequences: AAA81260 to AAA81303 and AAB25620 to AAB25663 represent

CC *Neisseria* DNA sequences and their corresponding proteins: AAA81254 to

CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the

CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to

CC AAA81452 represent *Neisseria meningitidis* MemB polynucleotide ORF

CC sequences, which are all used in the exemplification of the present

CC invention. The nucleic acid sequences, protein sequences, and antibodies

CC against them, can be used in the manufacture of a composition. The

CC composition can be used as a medicament (or in the manufacture of a

CC medicament) for treating, preventing or diagnosing infection due to

CC *Neisseria* bacteria. For example, some of the identified proteins could

CC be components of vaccines against *Meningococcus B*; against all serotypes;

CC and/or against all pathogenic *Neisseriae*. Identification of sequences

CC from the bacterium will also facilitate production of biological probes,

CC particularly organism-specific probes. Attempts to make effluents

CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.

CC Multivalent vaccines have also been tried but none have successfully

CC overcome antigenic variability. The provision of further, complete

CC sequences may provide an opportunity to identify secreted or surface

CC exposed proteins that may be presumed targets for the immune system and

CC which are not antigenically variable or at least more conserved than

CC other more variable regions.

XX Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

SQ

Query Match 71.2%; Score 17.8; DB 21; Length 1437668;

Best Local Similarity 90.5%; Pred. No. 54;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 catgccagcatctgcattcct 22

||||| ||||| ||||| |||||

DB 414663 catgccagcatctgcattcct 414683

RESULT 12

AAAX13229/C

ID AAAX13229 standard; DNA: 20561 BP.

XX

AC AAAX13229;

XX

DT 19-MAR-1999 (first entry)

XX

DE Enterococcus faecalis genome contig SEQ ID NO:292.

XX

KW Enterococcus faecalis; contig; detection; Enterococcal infection;

KW vaccine; attenuation; computer readable medium; ds.

XX

OS Enterococcus faecalis.

OS

PN WO9850555-A2.

XX

PD 12-NOV-1998.

XX

PF 04-MAY-1998; 98WO-0508985.

XX

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Barash SC, Dillon PJ, Kunsch CA;

XX WPI; 1999-045171/04.

DR

XX New isolated Enterococcus faecalis polynucleotides and polypeptides

PT - used to develop products for the detection of Enterococcus and for

PT use in vaccines for prevention or attenuation of Enterococcus

PT infection.

PS Claim 1; Page 1330-1340; 2084pp; English.

XX

XX A computer readable medium has been developed which has recorded on it

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC AAAX12938 to AAAX13919 represent these nucleotide sequences which are

CC primarily nucleotide sequences, also known as contigs. The computer-based

CC system can identify fragments of the Enterococcus faecalis genome with

CC commercial importance. The products can be used to detect the presence

CC of Enterococcus faecalis in samples. They can also be used for

CC diagnosing Enterococcal infection in an animal and monitoring

CC progression of disease, and for identifying agents which can be used to

CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC another related organism, in vivo or in vitro. In particular the

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC can be used in vaccines to prevent or attenuate an Enterococcal

CC infection.

XX Sequence 20561 BP; 6898 A; 3503 C; 4035 G; 6112 T; 13 other;

SQ

Query Match 70.4%; Score 17.6; DB 20; Length 20561;

Best Local Similarity 83.3%; Pred. No. 43;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 catgccagcatctgcattcct 25

||||| ||||| ||||| |||||

DB 9218 CATGCCAGCATCTTGCATCTCGT 9195

RESULT 13

AAAX99364

ID AAAX99364 standard; DNA: 6386 BP.

XX

AC AAAX99364;

XX

DT 09-NOV-1999 (first entry)

XX

DE Maize shrunken-1 (Sh1) gene sequence.

XX

KW Gene expression; shrunken-1 gene; Sh1 gene; maize; foreign gene; ds.

KW

XX Zea mays.

OS

XX

EH key

FT exon

FT 131..182

FT /*tag= a

FT /number= 1

FT 183..1210

FT /*tag= b

FT /number= 1

FT 1211..1324

FT /*tag= c

FT /number= 2

FT 1325..1827

FT /*tag= d

FT /number= 2

FT 1828..1948

FT /*tag= e

FT /number= 3

FT 1949..2040

FT /*tag= f

FT /number= 3

FT 2041..2187

FT /*tag= g

FT exon

```

FT      /number= 4
PT      2188..2268
PT      /tag= h
PT      /number= 4
PT      2269..2460
PT      /tag= i
PT      /number= 5
PT      2461..2604
PT      /tag= j
PT      /number= 5
PT      2605..2728
PT      /tag= k
PT      /number= 6
PT      2729..2821
PT      /tag= l
PT      /number= 6
PT      2822..3038
PT      /tag= m
PT      /number= 7
PT      3039..3255
PT      /tag= n
PT      /number= 7
PT      3256..3351
PT      /tag= o
PT      /number= 8
PT      3352..3446
PT      /tag= p
PT      /number= 8
PT      3447..3620
PT      /tag= q
PT      /number= 9
PT      3621..3701
PT      /tag= r
PT      /number= 9
PT      3702..3818
PT      /tag= s
PT      /number= 10
PT      3819..3911
PT      /tag= t
PT      /number= 10
PT      3912..4078
PT      /tag= u
PT      /number= 11
PT      4079..4157
PT      /tag= v
PT      /number= 11
PT      4158..4381
PT      /tag= w
PT      /number= 12
PT      4382..4516
PT      /tag= x
PT      /number= 12
PT      4517..4835
PT      /tag= y
PT      /number= 13
PT      4836..4767
PT      /tag= z
PT      /number= 13
PT      4768..5212
PT      /tag= aa
PT      /number= 14
PT      5213..5371
PT      /tag= ab
PT      /number= 14
PT      5372..5510
PT      /tag= ac
PT      /number= 15
PT      5511..5635
PT      /tag= ad
PT      /number= 15
PT      5636..5917
PT      /tag= ae
PT      /number= 16

```

```

XX      US955330-A.
PN      21-SEP-1999.
XX      /tag= i
XX      /number= 5
XX      2461..2604
XX      /tag= j
XX      /number= 5
XX      2605..2728
XX      /tag= k
XX      /number= 6
XX      2729..2821
XX      /tag= l
XX      /number= 6
XX      2822..3038
XX      /tag= m
XX      /number= 7
XX      3039..3255
XX      /tag= n
XX      /number= 7
XX      3256..3351
XX      /tag= o
XX      /number= 8
XX      3352..3446
XX      /tag= p
XX      /number= 8
XX      3447..3620
XX      /tag= q
XX      /number= 9
XX      3621..3701
XX      /tag= r
XX      /number= 9
XX      3702..3818
XX      /tag= s
XX      /number= 10
XX      3819..3911
XX      /tag= t
XX      /number= 10
XX      3912..4078
XX      /tag= u
XX      /number= 11
XX      4079..4157
XX      /tag= v
XX      /number= 11
XX      4158..4381
XX      /tag= w
XX      /number= 12
XX      4382..4516
XX      /tag= x
XX      /number= 12
XX      4517..4835
XX      /tag= y
XX      /number= 13
XX      4836..4767
XX      /tag= z
XX      /number= 13
XX      4768..5212
XX      /tag= aa
XX      /number= 14
XX      5213..5371
XX      /tag= ab
XX      /number= 14
XX      5372..5510
XX      /tag= ac
XX      /number= 15
XX      5511..5635
XX      /tag= ad
XX      /number= 15
XX      5636..5917
XX      /tag= ae
XX      /number= 16

Query Match          69.6%; Score 17.4; DB 20; Length 6386;
Best Local Similarity 94.7%; Pred. NO. 46;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 catgcacagcatcgcgc 20
DB      5698 catgcctgcctcgcgc 5716

RESULT 14
AAZ19606
ID      AAZ19606 standard; RNA: 61 BP.
XX
XX      AAZ19606;
AC
XX      08-NOV-1999 (first entry)
DT
XX
XX      Complement system protein C5 RNA binding ligand 13.
DE
XX
XX      Complement system protein; C1q; C5; C3; RNA ligand; nootropic; stroke;
KW      neuroprotective; anti-asthmatic; cerebroprotective; immunosuppressive;
KW      diagnostic; prevention; treatment; complement protein-related disease;
KW      Alzheimer's disease; renal disease; transplant rejection; asthma; ss.
XX
XX      Synthetic.
OS
XX
XX      WO941271-A1.
PN
XX
XX      19-AUG-1999.
PD
XX
XX      05-FEB-1999; 99WO-US02597.
PE
XX
XX      29-SEP-1998; 98US-0163025.
PR
XX      12-FEB-1998; 98US-0023228.
XX
XX      (NEXS-) NEXSTAR PHARM INC.
PA

```

Query Match	67.2%	Score 16.8;	DB 20;	Length 61;
Best Local Similarity	65.0%	Pred. No. 50;		
Matches	13;	Conservative	5;	Mismatches 2;
				Indels 0;
				Gaps 0
Qy	5	cgcgcagattcgatctctg	24	
		::: :		
db	20	cgcgcagcauucgcaucucg	39	

	CC	(1); expression products by culturing cells of (2);
	CC	(4) proteins or peptides with the same sequences (A); (5) poly- or mono-clonal antibodies (Ab) that react specifically with (A); (6)
	CC	hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic plants that contain transformed or transfected cells of (2); (8)
	CC	detecting KT2440 using a labeled (I) or Ab as probe; and (9) DNA chips carrying one or more (I), (1), and their fragments, are used as probes to detect and isolate full-length cDNAs and/or to amplify such cDNAs by polymerase chain reaction, and for production of transgenic plants. (1), or antibodies that recognize their expression products, are used for detecting the presence of KT2440, particularly in presence of other,
	CC	even closely related, bacteria. KT2440 is one of the bacteria classified as safe, by the National Institutes of Health, for genetic engineering work, e.g. as microbial production strains, for biological remediation and as vaccine carriers. (1) are exclusive to KT2440 with no significant homology with sequences in other bacteria (specifically those closely related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it has greater catalytic activity and better survival in, and adaptation to, the rhizosphere and soil.
SQ	XX	Sequence 1497 BP; 275 A; 515 C; 441 G; 264 T; 2 other;
	DY	Query Match Best Local Similarity 67.2%; Score 16.8; DB 22; Length 1497; Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0
Dn	Bd	cgcacgattcgatccgccg 24 cgcacgatgtcatcccgg 379
	RESULT 16	AAC34251/c
ID	XN	AAC34251 standard; DNA; 609 BP.
AC	XX	AAC34251:
DT	XX	17-OCT-2000 (first entry)
DE	XX	Arabidopsis thaliana DNA fragment SEQ ID NO: 5983.
KW	XX	Hybridisation assay; genetic mapping; gene expression control;
KM	XX	protein identification; signal transduction pathway;
KX	OS	metabolic pathway; promoter; termination sequence; ss. Arabidopsis thaliana. EP1033405-A2.
PB	PD	06-SEP-2000.
PF	PE	25-FEB-2000; 2000BP-0301439.
PR	PR	25-FEB-1999; 990S-0121825.
PR	PR	05-MAR-1999; 990S-0123180.
PR	PR	09-MAR-1999; 990S-0123548.
PR	PR	23-MAR-1999; 990S-0125788.
PR	PR	25-MAR-1999; 990S-0126264.
PR	PR	29-MAR-1999; 990S-0126785.
PR	PR	01-APR-1999; 990S-0127462.
PR	PR	06-APR-1999; 990S-0128234.
PR	PR	08-APR-1999; 990S-0128714.
PR	PR	16-APR-1999; 990S-0129845.
PR	PR	19-APR-1999; 990S-0130077.
PR	PR	21-APR-1999; 990S-0130449.
PR	PR	23-APR-1999; 990S-0130510.
PR	PR	28-APR-1999; 990S-0130891.
PR	PR	30-APR-1999; 990S-0131449.
RR	RR	30-APR-1999; 990S-0132048.
RR	RR	04-MAY-1999; 990S-0132407.
RR	RR	04-MAY-1999; 990S-0132484.

PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140699.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142055.
PR 02-JUL-1999; 99US-0142390.
PR 06-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142920.
PR 09-JUL-1999; 99US-0142927.
PR 12-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143624.
PR 14-JUL-1999; 99US-0144005.
PR 15-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.

PR 21-OCT-1999: 99US-0160770.
 PR 21-OCT-1999: 99US-0160814.
 PR 21-OCT-1999: 99US-0160815.
 PR 22-OCT-1999: 99US-0160980.
 PR 22-OCT-1999: 99US-0160981.
 PR 22-OCT-1999: 99US-0160989.
 PR 25-OCT-1999: 99US-0161404.
 PR 25-OCT-1999: 99US-0161405.
 PR 25-OCT-1999: 99US-0161406.
 PR 26-OCT-1999: 99US-0161359.
 PR 26-OCT-1999: 99US-0161360.
 PR 26-OCT-1999: 99US-0161361.
 PR 28-OCT-1999: 99US-0161920.
 PR 28-OCT-1999: 99US-0161922.
 PR 28-OCT-1999: 99US-0161993.
 PR 29-OCT-1999: 99US-0162142.

Query Match 66.4%; Score 16.6; DB 21; Length 609;
 Best Local Similarity 82.6%; Pred. No. 83;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcatcgccagcattcgatccgc 23
 ||||| ||||| ||||| ||
 DB 159 GCATCCACCGCATTCGATCATC 137

RESULT 17
 AAF1337/c
 ID AAF1337 standard; cDNA; 671 BP.

XX AAF13337;

DT 13-MAR-2001 (first entry)

DE Aspergillus oryzae EST SEQ ID NO:5860.

XX Multiple gene expression; filamentous fungal cell; EST;

KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; Identification; recombination;

KW culture condition; environmental stress; spore morphogenesis;

KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX Aspergillus oryzae.

OS WO200056762-A2.

XX 28-SEP-2000.

PD 22-MAR-2000: 2000MO-US07781.

PF 22-MAR-1999: 99US-0273623.

XX (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

PI WPI; 2000-594572/56.

DR Monitoring differential expression of genes in filamentous fungal cells

XX substrate of expressed sequence tags -

PT Claim 88; Page 2422-2423; 3161pp; English.

CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (PF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the PF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (PF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring

CC the global expression of genes from PF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how PF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF1247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.

SQ Sequence 671 BP; 164 A; 225 C; 157 G; 125 T; 0 other;

Query Match 66.4%; Score 16.6; DB 21; Length 671;
 Best Local Similarity 82.6%; Pred. No. 84;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 atcgccagcattcgatccgc 25
 ||||| ||||| ||||| ||
 DB 563 ACCGGCCGATTCGTCCTCCT 541

RESULT 18
 AAT79955/c
 ID AAT79955 standard; DNA; 501 BP.

XX AAT79955;

DT 16-OCT-1997 (first entry)

DE Fragment of cblIA gene of Pseudomonas cepacia.

XX Pseudomonas cepacia; transmissible lineage; Toronto/Edinburgh lineage;

KW cable adhesion type II PC pilli; pillin protein; cystic fibrosis;

KW cblA gene; ss.

XX Pseudomonas cepacia.

OS WO9701647-A2.

XX 16-JAN-1997.

PD 28-JUN-1996: 96MO-US11132.

PF 28-JUN-1995: 95US-0000828.

XX (HEAL-) HEALTH & HOSPITALS CITY BOSTON.

PI Goldstein R;

XX WPI; 1997-100217/09.

DR Identification of Pseudomonas cepacia lineages - using restriction

XX fragment length polymorphism analysis to identify highly

PT Transmissible strains

PS Disclosure; Fig 4; 52pp; English.

CC This sequence represents nucleotides 1-501 of the cblIA gene (encoding a
 CC 17 kDa major subunit pilin protein) of Pseudomonas cepacia (PC). This
 CC sequence is recognised by the primers and probes shown in
 CC AAT79942-T79954. PC is an aerobic gram-negative bacillus with a
 CC ubiquitous distribution in soil and water. It is an important pathogen
 CC among cystic fibrosis (CF) patients, where patients infected by PC have a
 CC higher morbidity and mortality than non-infected CF patients. The primers
 CC and probes that recognise this sequence can be used in the method of the

CC invention. The method of the invention is for detecting the presence of a
CC strain of a transmissible lineage of PC in a sample. The method comprises
CC analysing the sample for restriction fragment length polymorphisms
CC (RFLPs) linked to a strain known to be of a transmissible lineage of PC.
CC Alternatively, the method comprises using one or more pairs of
CC oligonucleotide primers (such as AAT79942-T79945) having sequences
CC identical to portions of the gene encoding a 17 kDa major subunit p11n
CC protein of the cable adhesin type II PC pill. The methods are used for
CC identifying highly transmissible lineages of PC, especially the
CC Toronto/Edinburgh lineage. They are used particularly for studying the
CC pathogen in CF patients.

SO Sequence 501 BP; 109 A; 154 C; 153 G; 85 T; 0 other;

Query Match 64.8%; Score 16.2; DB 18; Length 501;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 cgcgcagcatttcgactcctcgt 25
||| ||||| ||||| |||||
DB 114 CGACACGATTTCGACGCTGT 94

RESULT 19
AAC53704/C
ID AAC53704 standard; DNA; 543 BP.

XX AAC53704;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 75424.
DE
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 75424.
KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
XX
XX EPI033405-A2.
PM
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.

```

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

```

```

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Query Match          64.8%; Score 16.2; DB 21; Length 543;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

OY 3 atcgcagcattcgactc 23
DB 122 ATCGGCACATTTCGATCTTC 102

```

```

RESULT 20
AAFI2787/C
ID AAFI2787 standard; CDNA; 675 BP.

```

```

XX AAFI2787;

```

```

DT 13-MAR-2001 (first entry)
XX

```

```

DE Aspergillus oryzae EST SEQ ID NO:5310.
XX

```

```

KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombinant;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX

```

```

OS Aspergillus oryzae.
XX

```

```

PN WO200056762-A2.
XX

```

```

PD 28-SEP-2000.
XX

```

```

PF 22-MAR-2000; 2000MO-US07781.
XX

```

```

PR 22-MAR-1999; 99US-0273623.
XX

```

```

PA (NOVO ) NOVO NORDISK BIOTECH INC.
XX
PA (NOVO ) NOVO NORDISK AS.
XX

```

```

PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX

```

```

DR WPI; 2000-594572/56.
XX

```

```

PT Monitoring differential expression of genes in filamentous fungal cells
XX
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
XX
PT substrate of expressed sequence tags -
XX

```

```

PS Claim 88; Page 2215; 3161pp; English.
XX

```

```

CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an

```

CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC *Fusarium venenatum*; AAF11248 to AAF11853 represents ESTs from *Aspergillus*
CC *niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and
CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
CC all specifically claimed in the present invention.

XX
SQ Sequence 675 BP; 171 A; 183 C; 172 G; 143 T; 6 other;

Query Match 64.8%; Score 16.2; DB 21; Length 675;
Best Local Similarity 85.7%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

CY 1 gcatgcacgattcgatcc 21
|||||
Db 248 GCATCGCACCATTCGATGC 228

RESULT 21
ID AAC54459/c
XX AAC54459 standard; DNA: 1451 BP.
AC AAC54459;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 77945.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
XX
XX EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0131825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139751.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139869.
PR 23-JUN-1999; 99US-0140053.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145092.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155133.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159284.
PR 13-OCT-1999; 99US-0159285.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.

PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 64.8%; Score 16.2; DB 21; Length 1451;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 3 atcgcagattcagctc 23
|||||
Db 417 ATCGGCAACATTTCGATCTTC 397
RESULT 22
AAC54360/c
ID AAC54360 standard; DNA; 1452 BP.
XX
AC AAC54360;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 77596.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.

```
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 29-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145226.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.8%; Score 16.2; DB 21; Length 1452;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

OY 3 atcgccagcattgcactc 23
|||||
DB 417 ATCGGCACATTGCATCTC 397

RESULT 23
AAC39263/c
ID AAC39263 standard; DNA: 1546 BP.
XX AAC39263;
AC
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 23979.
DE
XX Hybridization assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158322.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match Best Local Similarity 64.8%; Score 16.2; DB 21; Length 1546;
Matches 18; Conservative 85.7%; Pred. No. 1.4e+02;

Mismatches 3; Indels 0; Gaps 0;

QY 3 atcgccagcattgcattc 23
|||||
DB 511 ATCGGCACATTCGATCTTC 491

RESULT 24

AAC54458/c
ID AAC54458 standard; DNA; 1549 BP.
XX
AAC54458;
AC
XX
18-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 77941.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 01-APR-1999; 99US-0126785.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137232.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.


```

DE Sorangium cellulosum 68.75 kb contig.
XX Epothilone biosynthesis; type I polyketide synthase; taxol substitute;
KM anticancer; ds.
XX Sorangium cellulosum.
XX Key Location/Qualifiers
FH 1..1826
FT /tag= a
FT /partial
FT /product= "Partial Orf 1 protein (AAV58580)"
FT /note= "No initiation codon given in the specification"
FT CDS complement (1900..3171)
FT
FT /tag= d
FT /product= "Orf 2 protein (AAV58581)"
FT /tag= c
FT /product= "Orf 3 protein (AAV58582)"
FT /product= "Orf 3 protein (AAV58582)"
FT /tag= d
FT /product= "Orf 4 protein (AAV58583)"
FT /tag= e
FT /product= "Orf 5 protein (AAV58584)"
FT CDS 7610..11875
FT /tag= f
FT /product= "Type I polyketide synthase, EPOS A
FT (AAV58573) " CDS 11872..116104
FT /tag= g
FT /product= "Non-ribosomal peptide synthetase, EPOS P
FT (AAV58574) "
FT CDS 16251..21749
FT /tag= h
FT /product= "Type I polyketide synthase, EPOS B
FT (AAV58575) " CDS 21746..43519
FT /tag= i
FT /product= "Type I polyketide synthase, EPOS C
FT (AAV58576) " CDS 43524..54920
FT /tag= j
FT /product= "Type I polyketide synthase, EPOS D
FT (AAV58577) " CDS 54935..62254
FT /tag= k
FT /product= "Type I polyketide synthase, EPOS E
FT (AAV58578) " CDS 62369..63628
FT /tag= l
FT /product= "Cytochrome P450 oxygenase homologue, EPOS F
FT (AAV58579) "
FT CDS 63779..64333
FT /tag= m
FT /product= "Orf 6 protein (AAV58585)"
FT /tag= n
FT /product= "Orf 7 protein (AAV58586)"
FT /tag= o
FT /product= "Orf 8 protein (AAV58587)"
FT CDS complement (64287..64727)
FT /tag= p
FT /product= "Orf 9 protein (AAV58588)"
FT CDS 65063..65767
FT /tag= q
FT /product= "Orf 10 protein"
FT /tag= r
FT /product= "Orf 11 protein (AAV58590)"
FT CDS complement (65871..66338)
FT /tag= s
FT /product= "Orf 12 protein (AAV58591)"
FT CDS 66667..67137
FT /tag= t
FT /product= "Orf 13 protein (AAV58592)"
FT CDS 67334..68251

```

```

FT /tag= u
FT /product= "Orf 14 protein (AAV58593)"
FT CDS 68346..68750
FT /tag= v
FT /product= "Partial Orf 15 protein (AAV58594)"
FT /note= "No termination codon given in the specification"
XX
XX MO996028-A2.
XX
XX 23-DEC-1999.
XX
XX 16-JUN-1999; 99WO-EP04171.
XX
XX 18-JUN-1998; 98US-0099504.
XX 24-SEP-1998; 98US-0101631.
XX 05-FEB-1999; 99US-0118906.
XX
XX (NOV ) NOVARTIS AG.
XX (NOV ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
XX Schupp T, Ligon JM, Molnar I, Zirkle R, Goerlach J, Cyr D;
XX WPT, 2000-097741/08.
XX P-PSDB; AAV58573, AAV58574, AAV58575, AAV58576, AAV58577, AAV58578,
XX AAV58579, AAV58580, AAV58581, AAV58582, AAV58583, AAV58584,
XX AAV58585, AAV58586, AAV58587, AAV58588, AAV58590, AAV58591,
XX AAV58592, AAV58593, AAV58594.
XX
XX New isolated epothilone synthase genes, used for the recombinant
XX production of epothilone for use in cancer therapy
XX
XX Claim 14; Page 87-104; 174pp; English.
XX
XX This sequence represents a 68.75 kb contig from Sorangium cellulosum
XX comprising 22 open reading frames (ORFs) and includes genes encoding
XX proteins involved in the biosynthesis of epothilones. Epothilones A and
XX B are 16-membered macrocyclic polyketides with an acylcysteine-derived
XX starter unit; polyketides being synthesised from two-carbon building
XX blocks, the beta-carbon of which always carries a keto group. Each round
XX of two-carbon addition is carried out by a complex of enzymes known as
XX the polyketide synthase in a manner similar to fatty acid biosynthesis.
XX EPOS A (AAV58573) and EPOS P (AAV58574) are involved in formation of
XX the thiazole ring formation of epothilones, and EPOS B, EPOS C, EPOS D
XX and EPOS E (AAV58575-58578) are involved in polyketide backbone
XX formation. EPO F (AAV58579) is an epothilone macrolactone oxidase, and
XX the proteins Orf 3 (AAV58582) and Orf14 (AAV58593) are thought to be
XX involved in transport. Epothilones mimic the biological activity of
XX taxol, and may be substituted for taxol in cancer chemotherapeutic
XX compositions. Epothilones exhibit a much lower drop in potency against a
XX multiply drug-resistant cell line compared with taxol, and are
XX considerably less efficiently exported from such cells by the multidrug
XX resistance protein (MDR, or P-glycoprotein). Despite the potential of
XX epothilones as anticancer agents, they are problematical to produce on a
XX large scale. Epothilones are too complex for industrial scale chemical
XX synthesis, and Sorangium cellulosum is difficult to ferment, producing
XX poor yields of epothilones. The nucleic acids of the invention may be
XX used for the recombinant production of epothilones in a heterologous host
XX that is more amenable to fermentation.
XX
XX Sequence 68750 BP; 9596 A; 22458 C; 25537 G; 11159 T; 0 other;

```

```

Query Match 64.8%; Score 16.2; DB 21; Length 68750;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Oy 5 cgcagcattcgatctcgt 25
Db 41953 cgcagcattcgatctcgt 41973

```

RESULT 26

```

AAA67164/C
ID AAA67164 standard; DNA; 205 BP.
AC
XX
XX
AAA67164:
XX
XX
31-OCT-2000 (first entry)
XX
XX
Pinus radiata 1,3-beta-D-glucanase DNA sequence SEQ ID NO:165.
DE
XX
Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KM plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant; ds.
XX
XX
Pinus radiata.
OS
XX
WO200022092-A2.
PN
XX
20-APR-2000.
PD
XX
08-OCT-1999; 99WO-NZ00169.
PF
XX
13-OCT-1998; 98US-0170862.
PR
XX
11-AUG-1999; 99US-0148426.
XX
XX
(GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX
Biosberg LN;
PI
XX
WPI; 2000-339328/29.
DR
XX
XX
New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
XX polysaccharide content, composition or structure of the plant
XX
XX
Claim 1; Page 118; 301pp; English.
PS
XX
The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
SQ Sequence 205 BP; 57 A; 55 C; 42 G; 51 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 205;
Best Local Similarity 79.2%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 catgcgcagcattcgatcctcgt 25
   |||| |||| |||| ||||
DB 137 CATCACACGCTTCTGTTCTCGT 114

RESULT 27
AAA67162/C
ID AAA67162 standard; DNA; 297 BP.
AC
XX
XX
AAA67162:
XX
XX
31-OCT-2000 (first entry)
XX
XX
Pinus radiata 1,3-beta-D-glucanase DNA sequence SEQ ID NO:163.
DE
XX
Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KW transgenic plant; ds.

```

```

KM plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant; ds.
XX
XX
Pinus radiata.
OS
XX
WO200022092-A2.
PN
XX
20-APR-2000.
PD
XX
08-OCT-1999; 99WO-NZ00169.
PF
XX
13-OCT-1998; 98US-0170862.
PR
XX
11-AUG-1999; 99US-0148426.
XX
XX
(GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX
Biosberg LN;
PI
XX
WPI; 2000-339328/29.
DR
XX
XX
New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
XX polysaccharide content, composition or structure of the plant
XX
XX
Claim 1; Page 117; 301pp; English.
PS
XX
The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
SQ Sequence 297 BP; 73 A; 80 C; 69 G; 75 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 297;
Best Local Similarity 79.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 catgcgcagcattcgatcctcgt 25
   |||| |||| |||| ||||
DB 243 CATCACACGCTTCTGTTCTCGT 220

RESULT 28
AAA67165/C
ID AAA67165 standard; DNA; 393 BP.
AC
XX
XX
AAA67165:
XX
XX
31-OCT-2000 (first entry)
XX
XX
Pinus radiata 1,3-beta-D-glucanase DNA sequence SEQ ID NO:166.
DE
XX
Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KM plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KW transgenic plant; ds.
XX
XX
Pinus radiata.
OS
XX
WO200022092-A2.
PN
XX
20-APR-2000.
PD
XX
08-OCT-1999; 99WO-NZ00169.

```

```

XX 13-OCT-1998; 98US-0170862.
PR 11-AUG-1999; 99US-0148426.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Bloksberg LN;
PI
XX WPI; 2000-339328/29.
XX
XX New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant
XX
XX Claim 1; Page 118; 301pp; English.
XX
XX The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AA16268
CC to AA16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
XX Sequence 393 BP; 95 A; 106 C; 91 G; 101 T; 0 other;
SQ

```

Query Match 64.0%; Score 16; DB 21; Length 393;
 Best Local Similarity 79.2%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 2 catgcgcagcattcgatcctcgt 25
    |||| |||| |||| |||| ||
Db 317 CATCACACGCTTTCGTCTCTCGT 294

```

RESULT 29
 AAA67166/c
 ID AAA67166 standard; DNA; 394 BP.
 XX
 AC AAA67166;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE Pinus radiata 1,3-beta-D-glucanase DNA sequence SEQ ID NO:167.
 XX
 KM Eucalyptus grandis; pinus radiata; Monterey pine; modification;
 KM plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
 KM transgenic plant; ds.
 XX
 OS Pinus radiata.
 XX
 WO200022092-A2.
 PN
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999; 99WO-NZ00169.
 XX
 PR 13-OCT-1998; 98US-0170862.
 PR 11-AUG-1999; 99US-0148426.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 XX Bloksberg LN;
 PI
 XX WPI; 2000-339328/29.
 DR

```

XX New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant
XX
XX Claim 1; Page 118; 301pp; English.
XX
XX The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AA16268
CC to AA16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
XX Sequence 394 BP; 95 A; 106 C; 91 G; 102 T; 0 other;
SQ

```

Query Match 64.0%; Score 16; DB 21; Length 394;
 Best Local Similarity 79.2%; Pred. No. 1.5e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 2 catgcgcagcattcgatcctcgt 25
    |||| |||| |||| |||| ||
Db 317 CATCACACGCTTTCGTCTCTCGT 294

```

RESULT 30
 AAA67357/c
 ID AAA67357 standard; DNA; 399 BP.
 XX
 AC AAA67357;
 XX
 DT 31-OCT-2000 (first entry)
 XX
 DE Pinus radiata endo-1,3-1,4-beta glucanase DNA sequence SEQ ID NO:358.
 XX
 KM Eucalyptus grandis; pinus radiata; Monterey pine; modification;
 KM plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
 KM transgenic plant; ds.
 XX
 OS Pinus radiata.
 XX
 WO200022092-A2.
 PN
 PD 20-APR-2000.
 XX
 PF 08-OCT-1999; 99WO-NZ00169.
 XX
 PR 13-OCT-1998; 98US-0170862.
 PR 11-AUG-1999; 99US-0148426.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 XX Bloksberg LN;
 PI
 XX WPI; 2000-339328/29.
 DR

New genes encoding proteins involved in a plant polysaccharide biosynthetic pathway, useful for modulating or altering the polysaccharide content, composition or structure of the plant

Claim 1; Page 166; 301pp; English.

The present invention describes isolated polynucleotides (PN) comprising a sequence selected from one of 835 nucleotide sequences given in AAA67073 to AAA67907, their (reverse) complements, sequences producing

CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them. 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
SQ Sequence 399 BP; 96 A; 108 C; 93 G; 102 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 399;
Best Local Similarity 79.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 catgccagcatttcgactcgt 25
|||||
DB 317 CATCACACAGCTTTCTGTCTCGT 294

RESULT 31
AAA67169/C
ID AAA67169 standard; DNA; 419 BP.
XX
AC AAA67169;

XX 31-OCT-2000 (first entry)

XX Pinus radiata 1,3-beta-D-glucanase DNA sequence SEQ ID NO:170.

XX Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KM plant cell wall; polysaccharide; polysaccharide biosynthetic
KM transgenic plant; ds.
XX
OS Pinus radiata.

PN WO200022092-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-NZ00169.

XX 13-OCT-1998; 98US-0170862.

PR 11-AUG-1999; 99US-0148426.

XX (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Bloksberg LN;

XX WPI; 2000-339328/29.

XX New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant -
XX
PS Claim 1; Page 119; 301pp; English.

XX The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX

SQ Sequence 419 BP; 104 A; 114 C; 98 G; 103 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 419;
Best Local Similarity 79.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 catgccagcatttcgactcgt 25
|||||
DB 317 CATCACACAGCTTTCTGTCTCGT 294

RESULT 32
AAA67358/C
ID AAA67358 standard; DNA; 469 BP.
XX
AC AAA67358;

XX 31-OCT-2000 (first entry)

XX Pinus radiata endo-1,3-1,4-beta glucanase DNA sequence SEQ ID NO:359.

XX Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KM plant cell wall; polysaccharide; polysaccharide biosynthetic
KM transgenic plant; ds.
XX
OS Pinus radiata.

PN WO200022092-A2.

XX 20-APR-2000.

XX 08-OCT-1999; 99WO-NZ00169.

XX 13-OCT-1998; 98US-0170862.

PR 11-AUG-1999; 99US-0148426.

XX (GENE-) GENESIS RES & DEV CORP LTD.

PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Bloksberg LN;

XX WPI; 2000-339328/29.

XX New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant -
XX
PS Claim 1; Page 166; 301pp; English.

XX The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX

SQ Sequence 469 BP; 119 A; 122 C; 105 G; 123 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 469;
Best Local Similarity 79.2%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 catgccagcatttcgactcgt 25
|||||
DB 317 CATCACACAGCTTTCTGTCTCGT 294

```

RESULT 33
AAA67167/c
ID AAA67167 standard; DNA: 498 BP.
XX
AC AAA67167;
XX
DT 31-OCT-2000 (first entry)
XX
DE Pinus radiata 1,3-beta-D-glucanase DNA sequence SEQ ID NO:168.
XX
KM Eucalyptus grandis; pinus radiata; Monterey pine; modification;
KM plant cell wall; polysaccharide; polysaccharide biosynthetic pathway;
KM transgenic plant; ds.
XX
OS Pinus radiata.
XX
PN WO200022092-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-N200169.
XX
PR 13-OCT-1998; 98US-0170862.
PR 11-AUG-1999; 99US-0148426.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bloksberg LN;
XX
DR WPI: 2000-339328/29.
XX
PT New genes encoding proteins involved in a plant polysaccharide
PT biosynthetic pathway, useful for modulating or altering the
PT polysaccharide content, composition or structure of the plant
XX
PS Claim 1; Page 119; 301pp; English.
XX
CC The present invention describes isolated polynucleotides (PN) comprising
CC a sequence selected from one of 835 nucleotide sequences given in
CC AAA67073 to AAA67907, their (reverse) complements, sequences producing
CC an Expectation (E) value of 0.01 or less compared to the 835 sequences,
CC sequences at least 50% identical to them, 200, 100, 40 or 20-mers of the
CC 835 sequences or sequences that are degenerately equivalent or allelic
CC to the 835 sequences. The polynucleotides are used to modify the
CC activity of a polypeptide involved in a polysaccharide biosynthetic
CC pathway in the plant. They are especially used to modulate or alter the
CC polysaccharide content, composition or structure of the plant. AAB16268
CC to AAB16340 are proteins encoded by some of the polynucleotide sequence
CC given in the present invention.
XX
SO Sequence 498 BP; 126 A; 129 C; 110 G; 130 T; 3 other:

Query Match 64.0%; Score 16; DB 21; Length 498;
Best Local Similarity 79.2%; Pred. No. 1.6e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 2 catgccagcatttcgacctcgt 25
Db 316 CATCACGAGCTTTCGTGTCGT 293

RESULT 34
AAF67973
ID AAF67973 standard; DNA: 822 BP.
XX
AC AAF67973;
XX
DT 11-APR-2001 (first entry)
XX

```

```

DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:461.
XX
KM Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KM membrane construction and membrane transport protein; petroleum spill;
KM hydrocarbon degradation; gram positive aerobic bacterium; marker;
KM identification; microorganism; fine chemical production; transformation;
KM genome mapping; genetic engineering; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100805-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-1B00926.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042088.
XX
PA (BAD1 ) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI: 2001-071486/08.
XX
PT P-PSDB: AAB76740.
XX
PT Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
PT transformation.
XX
PS Claim 3; Page 807-808; 1119pp; English.
XX
CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC construction and membrane transport (MCT) proteins given in AAB76510 to
CC AAB76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points

```


PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139470.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140685.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.

PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.0%; Score 16; DB 21; Length 1135;
Best Local Similarity 79.2%; Pred. No. 1.7e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gcatgcagcagcttcgctcgcg 24
||||| ||||| ||||| ||||| |||||
DB 426 GCATCACCAGCATTTTATCTCTG 403

RESULT 37

AAx80215
ID AAX80215 standard; CDNA: 1280 BP.

AAx80215;

18-AUG-1999 (first entry)

DE Mouse nuclear receptor protein nNR4 encoding cDNA.

KW Mouse; nuclear receptor protein; nNR4; identification; differentiation;
cell proliferation; regulation; murine; ss.

OS Mus musculus.

Key Location/Qualifiers

FT CDS 155..1228
FT /*tag= a

PN W09929722-A1.

PD 17-JUN-1999.

PF 11-DEC-1998; 98WO-US26446.

PR 12-DEC-1997; 97US-0068144.

PA (MERI) MERCK & CO INC.

PI Chen F;

DR WPI: 1999-385573/32.

DR P-PSDB; AAT17872.

PT Novel DNA encoding murine nuclear receptor

PS Claim 9; Page 30-31; 47pp; English.

CC The present sequence encodes a mouse nuclear receptor protein designated
CC nNR4. The nNR4 protein is useful in the identification of downstream
CC target genes and ligands regulating its activity. The nuclear receptor
CC is involved in the regulation of in vivo cell proliferation and/or cell
CC development. The nNR4 polynucleotides, expression vectors and host cells
CC are useful for the recombinant production of the protein.
XX
SQ Sequence 1280 BP; 309 A; 348 C; 343 G; 280 T; 0 other;

Query Match 64.0%; Score 16; DB 20; Length 1280;
Best Local Similarity 79.2%; Pred. No. 1.8e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gcatgcagcagcttcgctcgcg 24
||||| ||||| ||||| ||||| |||||
DB 613 gcatcacgcgcttcgctcgcg 636

RESULT 38

AAx24003
ID AAX24003 standard; DNA: 1361 BP.

AC AAX24003;
XX
XX 29-JUN-1999 (first entry)

DE Mouse CAR receptor DNA.

KW CAR receptor; constitutive activator of retinoic acid response element;
cytostatic; antithyroid; nuclear hormone receptor superfamily; cancer;
KW zinc finger transcription factor; retinoic acid; treatment; lung;
KW Grave's disease; thyroid hormone receptor; mouse; ss.

OS Mus sp.

Key Location/Qualifiers

FT CDS 148..1224
FT /*tag= a
FT /product= "CAR receptor"

PN W09915555-A1.

PD 01-APR-1999.

PF 17-SEP-1998; 98WO-US19365.

PR 19-SEP-1997; 97US-0934388.

PA (GENO) GEN HOSPITAL CORP.

PI Baes MI, Choi H, Moore DD;

DR WPI: 1999-254691/21.

DR P-PSDB; AAW93903.

PT CAR (Constitutive Activator of Retinoic Acid Response Elements)
PT receptor polypeptides, used to identify ligands for therapy of
PT Grave's disease and lung cancer

PS Claim 2; Fig 2; 66pp; English.

CC This invention describes the isolation of novel human and mouse CAR
CC (Constitutive Activator of Retinoic Acid Response Elements) receptor
CC polypeptides which have cytostatic and antithyroid activity. The CAR
CC receptor polypeptides are members of the nuclear hormone receptor
CC superfamily (zinc finger transcription factors). The CAR receptor
CC polypeptides bind to their target DNA sequence and activate expression of
CC downstream genes, even in the absence of retinoic acid. CAR receptor
CC ligands are useful for treating Grave's disease by decreasing thyroid
CC hormone receptor function, and for treating cancer (especially lung
CC cancer) by increasing retinoic acid receptor expression.
XX
SQ Sequence 1361 BP; 344 A; 362 C; 364 G; 291 T; 0 other;

Query Match 64.0%; Score 16; DB 20; Length 1361;
Best Local Similarity 79.2%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 gcatgcagcagcttcgctcgcg 24
||||| ||||| ||||| ||||| |||||
DB 606 gcatcacgcgcttcgctcgcg 629

RESULT 39

AAx72944
ID AAT72944 standard; CDNA: 2470 BP.

AAx72944;

10-MAR-1998 (first entry)

DE Phaffia carotenoid biosynthesis pathway enzyme gene Prcty.
XX
XX Glyceraldehyde-3-phosphate dehydrogenase; GAPDH; carotenoid;
KW

KW	synthesis; promoter; recombinant DNA; astaxanthin; ribosomal protein;
KW	Food colouring; ds.
XX	
OS	Phaffia rhodozyma.
XX	
FH	Key
FH	Location/Qualifiers
CDS	177..2198
FT	/*tag= a
FT	/product= Prcrty
XX	
PN	W09723633-A1.
XX	
PD	03-JUL-1997.
XX	
PF	23-DEC-1996; 96WO-EP05887.
XX	
PR	11-APR-1996; 96EP-0200943.
PR	22-DEC-1995; 95EP-0203620.
PA	(KONN) GIST-BROCADES BV.
PA	(OOIJ/) OOIJEN A J J.
PI	
PI	Verdoes JC, Wery J;
XX	
DR	WPI: 1997-351068/32.
DR	P-PSDB: AAW22500.
XX	
PT	Phaffia derived GAPDH and carotenoid synthesis genes and promoter
PT	fragment - used in the recombinant production of therapeutically
PT	useful proteins e.g. carotenoids for use in food colouring
XX	
PS	Claim 32: Page 60-63; 118pp: English.
XX	
CC	The present sequence represents a Phaffia derived carotenoid
CC	biosynthesis pathway enzyme gene which can be used in the novel
CC	recombinant DNA of the present invention. The recombinant DNA comprises
CC	a transcription promoter operably linked to a downstream sequence to be
CC	expressed, where the transcription promoter comprises a region found
CC	upstream of the open reading frame (ORF) of a highly expressed Phaffia
CC	gene (preferably GAPDH, a ribosomal protein gene, an enzyme involved in
CC	the biosynthesis pathway). The recombinant DNA can be used to transform
CC	hosts, preferably Phaffia. These transformed hosts are then used in the
CC	recombinant production of GAPDH or an enzyme involved in carotenoid
CC	synthesis, preferably astaxanthin. They may also be used to produce a
CC	pharmaceutical product. Purified carotenoids can be used as colourants
CC	in food and/or feed, and also in cosmetics.
XX	
CC	
SQ	Sequence 2470 BP; 605 A; 659 C; 531 G; 675 T; 0 other;
XX	
Query Match	64.0%; Score 16; DB 18; Length 2470;
Best Local Similarity	79.2%; Pred. No. 1.9e+02;
Matches 19; Conservative	0; Mismatches 5; Indels 0; Gaps 0
QY	2 catcgccagatctgcgacgtcgt 25
Db	275 catctacaaatcatgcgtcgtcgt 298
XX	
RESULT 40	
AAAT72941	
ID 25	AAAT72941 standard; cDNA: 2546 BP.
XX	
AC	AAAT72941;
XX	
DT	10-MAR-1998 (first entry)
XX	
DE	Phaffia carotenoid biosynthesis pathway enzyme gene PrcrTB.
XX	
KW	Glyceraldehyde-3-phosphate dehydrogenase; GAPDH; carotenoid;
KW	synthesis; promoter; recombinant DNA; astaxanthin; ribosomal
KW	protein; food colouring; ds.
XX	

OS	Phaffia rhodozyma.
FH	Key
FT	Location/Qualifiers
CD	225..2246
CDS	/tag= a
ET	/product= pRCrTB
XX	
PN	M09723633-A1.
XX	
PD	03-JUL-1997.
XX	
PF	23-DEC-1996; 96WO-EP0587.
XX	
PR	11-APR-1996; 96EP-0200943.
PR	22-DEC-1995; 95EP-0203620.
XX	
PA	(KONN) GIST-BROCADES BV.
PA	(OOIJ/) OOIJEN A J J.
XX	
PI	Verdoes JC, Wery J;
PI	
DR	WPI: 1997-351068/32.
DR	P-PSDB; AAM22497.
XX	
PT	Phaffia derived GAPDH and carotenoid synthesis genes and promoter
PT	fragment - used in the recombinant production of therapeutically
PT	useful proteins e.g. carotenoids for use in food colouring
XX	
PS	Claim 32; Page 48-50; 118pp; English.
XX	
CC	The present sequence represents a Phaffia derived carotenoid
CC	biosynthesis pathway enzyme gene which can be used in the novel
CC	recombinant DNA of the present invention. The recombinant DNA comprises
CC	a transcription promoter operably linked to a downstream sequence to be
CC	expressed, where the transcription promoter comprises a region found
CC	upstream of the open reading frame (ORF) of a highly expressed Phaffia
CC	gene (preferably GAPDH, a ribosomal protein gene, an enzyme involved in
CC	the biosynthesis pathway). The recombinant DNA can be used to transform
CC	hosts, preferably Phaffia. These transformed hosts are then used in the
CC	recombinant production of GAPDH or an enzyme involved in carotenoid
CC	synthesis, preferably astaxanthin. They may also be used to produce a
CC	pharmaceutical product. Purified carotenoids can be used as colourants
CC	in food and/or feed, and also in cosmetics.
XX	
SO	Sequence 2546 BP; 628 A; 676 C; 552 G; 690 T; 0 other;
Query Match	64.0%; Score 16; DB 18; Length 2546;
Best Local Similarity	79.2%; Pred.No. 1.9e+02;
Matches 19; Conservative	0; Mismatches 5; Indels 0; Gaps 0;
OY	2 catcgccagcattgcgtctcgt 25
Dd	323 cattcaaaaatcatgctcgtcgt 346
RESULT 41	
AAT72946	ID AAT72946 Standard; DNA; 3550 BP.
XX	
AC	AAT72946;
XX	
DT	10-MAR-1998 (first entry)
XX	
DE	Phaffia carotenoid biosynthesis pathway enzyme gene pRCrTB GB.
KM	Glyceraldehyde-3-phosphate dehydrogenase; GAPDH; carotenoid;
KM	synthesis; promoter; recombinant DNA; astaxanthin; protein;
KM	food colouring; ds.
XX	
SS	Phaffia rhodozyma.
XX	
HH	Location/Qualifiers

FT	CDS	941..3325
FT		/tag= a
FT		/product= PRcGrLB_GB
FT	exon	/note= "contains introns"
FT		941..966
FT		/tag= b
FT		/number= 1
FT	intron	967..1077
FT		/tag= c
FT		/number= 1
FT	exon	1078..1284
FT		/tag= d
FT		/number= 2
FT	intron	1285..1364
FT		/tag= e
FT		/number= 2
FT	exon	1365..1877
FT		/tag= f
FT		/number= 3
FT	intron	1878..1959
FT		/tag= g
FT		/number= 3
FT	exon	1960..2202
FT		/tag= h
FT		/number= 4
FT	intron	2203..2292
FT		/tag= i
FT		/number= 4
FT	exon	2293..3325
FT		/tag= j
FT		/number= 5
XX		
PN		W09723633-A1.
PD		03-JUL-1997.
XX		
PF	23-DEC-1996;	96WO-EP05887.
XX		
PR	11-APR-1996;	96EP-0200943.
PR	22-DEC-1995;	95EP-0203620.
XX		
PA	(KONN) GIST-BROCADES BV.	
PA	(OOLJ/) OOIJEN A J J.	
XX		
PI	Verdoes JC, Wery J;	
XX		
DR	WPI; 1997-351068/32.	
DR	P-PSDB; AAW22502.	
XX		
PT	Phaffia derived GAPDH and carotenoid synthesis genes and promoter	
PT	fragment - used in the recombinant production of therapeutically	
PT	useful proteins e.g. carotenoids for use in food colouring	
XX		
PS	Claim 32; Page 68-72; 118pp; English.	
XX		
CC	The present sequence represents a Phaffia derived carotenoid	
CC	biosynthesis pathway enzyme gene which can be used in the novel	
CC	recombinant DNA of the present invention. The recombinant DNA comprises	
CC	a transcription promoter operably linked to a downstream sequence to be	
CC	expressed, where the transcription promoter comprises a region found	
CC	upstream of the open reading frame (ORF) of a highly expressed Phaffia	
CC	gene (preferably GAPDH, a ribosomal protein gene, an enzyme involved in	
CC	the biosynthesis pathway). The recombinant DNA can be used to transform	
CC	hosts, preferably Phaffia. These transformed hosts are then used in the	
CC	recombinant production of GAPDH or an enzyme involved in carotenoid	
CC	synthesis, preferably astaxanthin. They may also be used to produce a	
CC	pharmaceutical product. Purified carotenoids can be used as colourants	
CC	in food and/or feed, and also in cosmetics.	
XX		
SQ	Sequence 3550 BP; 840 A; 926 C; 759 G; 1025 T; 0 other;	

[illegible]

```

FT CDS 4193..4371
FT /*tag= m
FT /label= CLR
FT /note= "putative CLR gene"
XX
XX
XX WO9737031-A1.
XX
XX
XX 09-OCT-1997.
XX
XX
XX 27-MAR-1997; 97WO-NZ00040.
XX
XX
XX 29-MAR-1996; 96NZ-0286284.
XX
XX
XX (UYOT-) UNIV OTAGO.
XX
XX
XX Lytle DJ, Robinson AJ;
XX
XX WPI; 1997-503116/46.
XX
XX Parapoxvirus vectors containing exogenous DNA - comprise parapox
XX virus, especially orf virus, and exogenous DNA encoding a
XX polypeptide, e.g. an antigen useful in vaccines
XX
XX
XX Claim 12; Fig 7; 73pp; English.
XX
XX This genomic DNA molecule comprises the BamHI F fragment and part
XX of the BamHI C fragment of the orf virus strain NZ-2 genome
XX incorporating putative promoter sequences for the FL1, F2L, F4R
XX and CLR genes as well as intergenic regions suitable as foreign
XX gene insertion sites. A claimed parapoxvirus vector comprises a
XX vector fragment (especially an orf virus) containing exogenous DNA.
XX and host cells (especially bovine or ovine testis cells)
XX incorporating the vector are also claimed. The vector and
XX fragments/variants are useful in vaccine production (claimed),
XX since they can express antigens, useful in medical and veterinary
XX applications e.g. parasitic disease control in livestock. Particular
XX antigens include HIV envelope protein, herpes simplex glycoprotein,
XX antigens of Echinococcus granulosus, Trichostrongylus, Haemonchus,
XX Ostertagia and especially Taenia ovis 45W, 16 kd, 18 kd antigens.
XX The foreign gene is preferably incorporated in a non-essential
XX region of the virus genome (see AAT91357-61) under control of a
XX poxvirus promoter, especially orf virus promoter FL1, FL1 or F3R
XX (see AAT91362-64). The orf virus vectors limit the cross-species
XX infection common with e.g. vaccinia virus vectors, since the virus
XX has a narrow host range; orf is also less virulent than vaccinia in
XX man.
XX
XX Sequence 4372 BP; 817 A; 1346 C; 1439 G; 770 T; 0 other;
SQ

```

Query Match 64.0%; Score 16; DB 18; Length 4372;
 Best Local Similarity 79.2%; Pred. No. 2e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

OY 1 gcatgcgacattgcattcctcg 24
   ||| |||| |||| |||| ||||
Db 3252 GCAGCGCCGCAATCCGAGCCTCG 3229

```

RESULT 43
 AAQ15312/c
 ID AAQ15312 standard; DNA: 7475 BP.
 XX
 XX AAQ15312;
 XX
 XX 17-MAR-1992 (first entry)
 XX
 XX Plasmid pPB12 encoding enzyme able to transform 5-substd. hydantoin
 XX into N-carbamyl amino acid.
 XX
 XX Hydrolyase; ss.
 XX

```

OS Pseudomonas NS 671 (FERM P-9543).
XX
XX Key Location/Qualifiers
XX CDS 1156..3228
XX
XX /*tag= a
XX /label= ORF 2
XX
XX CDS 3232..5010
XX
XX /*tag= b
XX /label= ORF 3
XX
XX CDS 5031..6275
XX
XX /*tag= c
XX /label= ORF 4
XX
XX
XX JP03251176-A.
XX
XX
XX 08-NOV-1991.
XX
XX
XX 29-MAY-1990; 90JP-0137120.
XX
XX
XX 10-JAN-1990; 90JP-0002786.
XX
XX 29-MAY-1990; 90JP-0137120.
XX
XX (NIPS ) NIPPON SODA KK.
XX
XX
XX WPI; 1991-373414/51.
XX
XX DR P-PSDB; AAR15482, AAR15483, AAR15484.
XX
XX Enzyme for transforming 5-substd. hydantoin - in which enzymatic
XX protein has specified aminoacid sequence.
XX
XX Disclosure; Fig 3; 11pp; Japanese.
XX
XX
XX The sequence comprises three open reading frames: 2, 3, and 4,
XX which together encode an enzyme which has the ability to transform
XX 5-substd. hydantoin into N-carbamyl amino acids. The DNA segment
XX contg. the three ORFs was obt'd. from Pseudomonas NS 671 and
XX ligated into pUC18 to form the plasmid pPB12.
XX
XX Sequence 7475 BP; 2424 A; 1237 C; 1698 G; 2116 T; 0 other;
SQ

```

Query Match 64.0%; Score 16; DB 12; Length 7475;
 Best Local Similarity 79.2%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

OY 1 gcatgcgacattgcattcctcg 24
   ||||| ||||| ||||| |||
Db 2992 GGATCGGCTGCTTTCGATTTTCG 2969

```

RESULT 44
 AAX03047
 ID AAX03047 standard; DNA: 9158 BP.
 XX
 XX AAX03047;
 XX
 XX 22-JUN-1999 (first entry)
 XX
 XX Human IL-1ra BAC contiguous DNA sequence 92.
 XX
 XX Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
 XX interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
 XX chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
 XX growth factors; treatment; IL-1 receptor complex; BAC; ss.
 XX
 XX Homo sapiens.
 XX
 XX WO9906426-A1.
 XX
 XX 11-FEB-1999.
 XX
 XX 03-AUG-1998; 98WO-US16102.
 XX

PR 02-JUL-1998: 9805-0091650.
 XX 04-AUG-1997: 9705-0054646.
 XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 XX Pan Y:
 XX WPI: 1999-153692/13.
 XX
 PT New isolated nucleic acid encoding the new human cytokine Tango-77 -
 XX used to inhibit inflammation and to screen for specific modulators
 XX
 PS Example 5; Figure 4; 226bp; English.
 XX
 CC AAX02956-X03048 and AAX22301-X22304 are overlapping BAC genomic
 CC sequences containing alternatively spliced forms of human IL-1ra. Such
 CC fragments are used in the method of the invention which describes the
 CC isolation of a novel human TANGO-77 encoding nucleic acid and protein.
 CC Tango-77 is a member of the cytokine superfamily that is expected to
 CC inhibit inflammation by binding to the interleukin-1 receptor (IL-1R). It
 CC may also bind to a new receptor so could regulate other cellular
 CC processes associated with acute or chronic inflammation, e.g. asthma,
 CC chronic myelogenous leukemia, rheumatoid arthritis, psoriasis and
 CC inflammatory bowel disease. It may also induce or suppress interleukins,
 CC cytokines and growth factors. Modulators of this protein are used to
 CC treat or prevent conditions associated with abnormal levels of
 CC inflammation, or activity of IL-1 or its receptor complex.
 XX
 SQ Sequence 9158 BP; 2430 A; 2049 C; 1917 G; 2752 T; 10 other;
 XX
 Query Match 64.0%; Score 16; DB 20; Length 9158;
 Best Local Similarity 79.2%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 catccagcattcgcctcgt 25
 ||||| ||||| ||||| |||||
 DB 4330 catccagcattcgtctgt 4353
 ||||| ||||| ||||| |||||
 RESULT 45
 AAV58939
 ID AAV58939 standard; DNA; 9960 BP.
 XX
 AC AAV58939;
 XX
 DT 15-FEB-1999 (first entry)
 XX
 DE Mycobacterium smegmatis embCAB operon.
 XX
 KM Drug resistance; antibiotic resistance; antimycobacterial;
 KM ethambutol; embCAB operon; infection; vaccine; therapy; ds.
 XX
 OS Mycobacterium smegmatis.
 XX
 PN WO9841533-A1.
 XX
 PD 24-SEP-1998.
 XX
 PE 16-MAR-1998: 98WO-US05128.
 XX
 PR 20-MAR-1997: 97US-0822586.
 XX
 PA (YESH) UNIV YESHIVA EINSTEIN COLLEGE.
 XX
 PI Jacobs WR, Musser J, Telenti A;
 XX
 DR WPI: 1998-521160/44.
 DR P-PSDB: AAW73055-57.
 XX
 PT Wild type and mutated sequences of Mycobacterium embCAB operon -
 PT useful to e.g. identify ethambutol-resistant mycobacterial strains
 PT and produce antisense sequences to treat mycobacterial infections

XX
 PS Disclosure: Fig 6A-I; 62pp; English.
 XX
 CC This is the DNA sequence of the Mycobacterium smegmatis embCAB
 CC operon, which determines resistance to the antimycobacterial drug
 CC ethambutol (EMB). It includes the embA, embC and embB genes that
 CC encode proteins (see AAW73052-54) which are the target of action of
 CC M. smegmatis for EMB. To identify genes conferring EMB resistance,
 CC a genomic library from a high level EMB-resistant mutant of M.
 CC smegmatis was introduced into wild-type M. smegmatis mc2155. Four
 CC overlapping cosmids were identified which conferred a resistant
 CC phenotype. The minimum size fragment capable of conferring EMB
 CC resistance was 9 kb (PIMW99). PIMW99 plus 7 kb upstream M.
 CC smegmatis sequence was sequenced revealing 3 homologous open
 CC reading frames (embC, embB, embA) and 4 additional potential coding
 CC regions. Wild-type and mutated embCAB nucleic acid sequences are
 CC useful as probes used in the diagnosis of drug-resistant
 CC mycobacteria or to determine the susceptibility of mycobacteria to
 CC EMB. The nucleic acids are also useful in the treatment of
 CC mycobacterial infections; anti-DNA or anti-RNA sequences can be
 CC administered to inhibit embCAB operon mRNA activity (claimed). The
 CC invention additionally provides for the use of embCAM operon
 CC nucleic acid sequences as vaccines, or to improve existing
 CC vaccines.
 XX
 SQ Sequence 9960 BP; 1389 A; 3592 C; 3360 G; 1618 T; 1 other;
 XX
 Query Match 64.0%; Score 16; DB 19; Length 9960;
 Best Local Similarity 79.2%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 gcatccagcattcgcctcgt 24
 ||||| ||||| ||||| |||||
 DB 1784 gcatccagcattcgtcctcgt 1807
 ||||| ||||| ||||| |||||

Search completed: October 9, 2001, 11:39:28
 Job time: 1863 sec

Wed Oct 10 07:44:48 2001

us-09-396-196f-7.std.rng

Page 32

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:39:49 ; Search time 221.13 Seconds
(without alignments)
21.403 Million cell updates/sec

Title: US-09-396-196f-7
Perfect score: 25
Sequence: 1 gcatcgccagcattcgcctcgt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
4: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
5: /cgn2_6/prodata/2/ina/5A.COMB.seq:*
6: /cgn2_6/prodata/2/ina/5A.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2 US-08-401-068-7	Sequence 7, Appl1
2	25	100.0	1041	2 US-08-846-338-7	Sequence 7, Appl1
3	25	100.0	5872	3 US-08-411-768B-1	Sequence 1, Appl1
4	25	100.0	5872	3 US-08-411-768B-6	Sequence 6, Appl1
5	17.4	69.6	6386	2 US-08-483-376-1	Sequence 1, Appl1
6	16.8	67.2	61	3 US-09-023-228B-59	Sequence 59, Appl1
7	16.2	64.8	851	4 US-08-998-416-294	Sequence 294, App
8	16.2	64.8	12732	4 US-09-060-756-1	Sequence 1, Appl1
9	16.2	64.8	68750	3 US-09-335-409-1	Sequence 1, Appl1
10	16	64.0	9960	3 US-08-822-586-46	Sequence 46, Appl1
11	16	64.0	10095	3 US-08-822-586-45	Sequence 45, Appl1
12	16	64.0	176373	3 US-09-128-155-17	Sequence 17, Appl1
13	15.6	62.4	978	3 US-08-706-281A-17	Sequence 17, Appl1
14	15.6	62.4	1147	1 US-08-417-103-15	Sequence 15, Appl1
15	15.6	62.4	1351	1 US-07-816-283-5	Sequence 5, Appl1
16	15.6	62.4	1351	1 US-08-417-103-5	Sequence 5, Appl1
17	15.4	61.6	723	4 US-08-998-416-617	Sequence 617, App
18	15.4	61.6	1126	3 US-08-479-733A-27	Sequence 27, Appl1
19	15.4	61.6	1126	3 US-08-487-427-27	Sequence 27, Appl1
20	15.4	61.6	1126	3 US-08-479-727A-27	Sequence 27, Appl1
21	15.4	61.6	1126	3 US-08-482-369A-27	Sequence 27, Appl1
22	15.4	61.6	1126	5 PCT-US95-07439-27	Sequence 27, Appl1
23	15.4	61.6	1500	1 US-08-487-037-4	Sequence 4, Appl1
24	15.2	60.8	1712	2 US-08-632-598-1	Sequence 1, Appl1
25	15.2	60.8	1956	2 US-08-762-308-10	Sequence 10, Appl1
26	15.2	60.8	2099	1 US-08-299-849B-25	Sequence 25, Appl1
27	15.2	60.8	2099	2 US-08-142-368A-25	Sequence 25, Appl1

c	28	15.2	60.8	2099	3 US-08-967-727-25	Sequence 25, Appl1
c	29	15.2	60.8	2099	4 US-08-037-230D-25	Sequence 25, Appl1
c	30	15.2	60.8	2150	1 US-08-299-849B-24	Sequence 24, Appl1
c	31	15.2	60.8	2150	2 US-08-142-368A-24	Sequence 24, Appl1
c	32	15.2	60.8	2150	3 US-08-967-727-24	Sequence 24, Appl1
c	33	15.2	60.8	2150	4 US-08-037-230D-24	Sequence 24, Appl1
c	34	15.2	60.8	3029	1 US-08-693-214-1	Sequence 1, Appl1
c	35	15.2	60.8	4724	1 US-08-404-665-3	Sequence 3, Appl1
c	36	15.2	60.8	4724	1 US-08-404-671-3	Sequence 3, Appl1
c	37	15.2	60.8	4724	1 US-08-404-781-3	Sequence 3, Appl1
c	38	15.2	60.8	87350	3 US-08-781-891-79	Sequence 79, Appl1
c	39	15	60.0	102	1 US-08-247-475-24	Sequence 24, Appl1
c	40	15	60.0	102	1 US-08-479-650-24	Sequence 24, Appl1
c	41	15	60.0	102	1 US-08-191-866D-30	Sequence 30, Appl1
c	42	15	60.0	102	1 US-08-674-169-24	Sequence 24, Appl1
c	43	15	60.0	102	2 US-08-185-949B-30	Sequence 30, Appl1
c	44	15	60.0	918	1 US-08-328-322-3	Sequence 3, Appl1
c	45	15	60.0	1222	1 US-08-328-322-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-401-068-7
Sequence 7, Application US/08401068
Patent No. 5859335
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/351,970
FILING DATE: 08-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ. ID NO. 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcatttcgattccctcgt 25
|||||
Db 90 GCATGCCAGCATTTCGATTCCTCCTG 114

RESULT 2
US-08-846-338-7
Sequence 7, Application US/08846338
Patent No. 5869719
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 5869719atlis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/846,338
FILING DATE:
CLASSIFICATION: 80D
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SRO ID NO: 7;
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0057;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcatttcgattccctcgt 25
|||||
Db 90 GCATGCCAGCATTTCGATTCCTCCTG 114

RESULT 3
US-08-411-768B-1
Sequence 1, Application US/08411768B
Patent No. 6083712
GENERAL INFORMATION:
APPLICANT: Olwen Birch
APPLICANT: Johann Birch
APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SRO ID NO: 1;
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A-15/9
FEATURE:
NAME/KEY: CDS
LOCATION: 117..1157
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 117
OTHER INFORMATION: /product= "Biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
FEATURE:
NAME/KEY: CDS
LOCATION: 2295..3050
OTHER INFORMATION: /codon_start= 2295
OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "bioc"
OTHER INFORMATION: /number= 3
FEATURE:
NAME/KEY: CDS
LOCATION: 3750..5039
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3750
OTHER INFORMATION: /EC_number= 2.6.1.62
OTHER INFORMATION: /product= "DAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioA"
OTHER INFORMATION: /number= 5
OTHER INFORMATION: /standard_name=
OTHER INFORMATION: "S-adenosyl-L-methionine:8-amino-7-oxononanoate
OTHER INFORMATION: aminotransf."


```

? LOCATION: 5098..5574
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start= 5098
? OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
? OTHER INFORMATION: /product= "protein"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ORF1"
? OTHER INFORMATION: /number= 6
? FEATURE:
? NAME/KEY: -10_signal
? LOCATION: 45..49
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /standard_name= "promoter plac"
? FEATURE:
? NAME/KEY: -35_signal
? LOCATION: 23..28
? OTHER INFORMATION: /standard_name= "promoter plac"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 105..119
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /standard_name= "biob RBS no.9"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 2284..2297
? OTHER INFORMATION: /standard_name= "bioc RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 3742..3752
? OTHER INFORMATION: /standard_name= "bioa RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 5088..5100
? OTHER INFORMATION: /standard_name= "ORF1 RBS"
? FEATURE:
? NAME/KEY: terminator
? LOCATION: 5583..5644
? OTHER INFORMATION: /standard_name= "rho-independent
? OTHER INFORMATION: /transcriptional terminator"
? FEATURE:
? NAME/KEY: stem_loop
? LOCATION: 5583..5605
? FEATURE:
? NAME/KEY: promoter
? LOCATION: 1..96
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /function= "promoter plac"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /number= 2
? DOCUMENT NUMBER: WO 87/01391 B1
? FILING DATE: 26-Aug-1986
? PUBLICATION DATE: 07-Apr-1993
? US-08-411-768B-1

Query Match          100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0068;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcatgcacgacattcgcattcgcatt 25
    |||||||||||||||||||||||||
Db 206 GCATGCCAGCATTCGATTCCTCTCT 230

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
;

APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: PB030A15-9
FEATURE:
? NAME/KEY: CDS
? LOCATION: 1154..2308
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start= 1154
? OTHER INFORMATION: /EC_number= 2.3.1.47
? OTHER INFORMATION: /product= "KAPA synthase"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "bioF"
? OTHER INFORMATION: /number= 2
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 3043..3753
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start= 3043
? OTHER INFORMATION: /EC_number= 6.3.3.3
? OTHER INFORMATION: /product= "DPA synthase"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "bioP"
? OTHER INFORMATION: /number= 4
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 1141..1156
? OTHER INFORMATION: /standard_name= "bioF RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 3030..3045
? OTHER INFORMATION: /standard_name= "biOD RBS"
? OTHER INFORMATION: /number= 2
? DOCUMENT NUMBER: WO 87/01391 B1
```


STREET: 8400 E. Prentice Place #200
CITY: Denver
STATE: Colorado
COUNTRY: US
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,228B
FILING DATE: 12-FEBRUARY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01739
FILING DATE: 30 JAN 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/595,335
FILING DATE: 1 FEB 1996
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson, Esq.
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX50/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
FEATURE:
OTHER INFORMATION: All pyrimidines are 2'-F modified
US-09-023-228B-59

Query Match 67.2%; Score 16.8; DB 3; Length 61;
Best Local Similarity 65.0%; Pred. No. 24;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 5 cgcgcagattcgatccgcg 24
DB 20 CGCAGCAUUCGACUUCG 39

RESULT 7
US-08-998-416-294
Sequence 294: Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippesen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reblschung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSPYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Park
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 294:
SEQUENCE CHARACTERISTICS:
LENGTH: 851 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1244RP
US-08-998-416-294

Query Match 64.8%; Score 16.2; DB 4; Length 851;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 cgcgcagattcgatccgcg 25
DB 503 CGCAGCATTTGTCATCGT 523

RESULT 8
US-09-060-756-1
Sequence 1: Application US/09060756
Patent No. 6183957
GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchleser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 12732
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-09-060-756-1

Query Match 64.8%; Score 16.2; DB 4; Length 12732;
Best Local Similarity 85.7%; Pred. No. 79;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcatcgccagattcgatcc 21
DB 11641 gcagcgccagattcgatcc 11661

RESULT 9
US-09-335-409-1
Sequence 1: Application US/09335409
Patent No. 6121029

GENERAL INFORMATION:
APPLICANT: Schupp, Thomas
APPLICANT: Ligon, James
APPLICANT: Molnar, Istvan
APPLICANT: Zirkle, Ross
APPLICANT: Cyr, Devon
APPLICANT: Goerlach, Joern
TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
FILE REFERENCE: 4-30582A
CURRENT APPLICATION NUMBER: US/09/335,409
CURRENT FILING DATE: 1999-06-17
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 1
LENGTH: 68750
TYPE: DNA
ORGANISM: Sorangium cellulosum
US-09-335-409-1

Query Match 64.8%; Score 16.2; DB 3; Length 68750;
Best Local Similarity 85.7%; Pred. No. 94;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 cgcagcattcgcctcgt 25
DB 41953 cgcagccttcgcacctcgt 41973

RESULT 10
US-08-822-586-46
Sequence 46 Application US/08822586
Patent No. 6015890
GENERAL INFORMATION:
APPLICANT: WILLIAM R. JACOBS, JR., JAMES M. MUSSER AND
APPLICANT: AMALIO TELENTE
TITLE OF INVENTION: AN EMBGAB OPERON OF MYCOBACTERIA AND
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
MEDIUM TYPE: DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,586
FILING DATE: MARCH 20, 1997
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/437
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 9960
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
US-08-822-586-46

Query Match 64.0%; Score 16; DB 3; Length 9960;
Best Local Similarity 79.2%; Pred. No. 95;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcacgcagattcgcctcgt 24
DB 1784 GCATCAGCATCATCTCTCTCTCG 1807

RESULT 11
US-08-822-586-45
Sequence 45 Application US/08822586
Patent No. 6015890
GENERAL INFORMATION:
APPLICANT: WILLIAM R. JACOBS, JR., JAMES M. MUSSER AND
APPLICANT: AMALIO TELENTE
TITLE OF INVENTION: AN EMBGAB OPERON OF MYCOBACTERIA AND
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
STREET: 90 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10016
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 INCH 1.44 MB STORAGE
MEDIUM TYPE: DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/822,586
FILING DATE: MARCH 20, 1997
ATTORNEY/AGENT INFORMATION:
NAME: ELIZABETH A. BOGOSIAN
REGISTRATION NUMBER: 39,911
REFERENCE/DOCKET NUMBER: 96700/437
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 10095
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
US-08-822-586-45

Query Match 64.0%; Score 16; DB 3; Length 10095;
Best Local Similarity 79.2%; Pred. No. 96;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcacgcagattcgcctcgt 24
DB 1859 GCATCAGCATCATCTCTCTCTCG 1882

RESULT 12
US-09-128-155-17
Sequence 17 Application US/09128155
Patent No. 6117654
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001

;; CURRENT APPLICATION NUMBER: US/09/128,155
;; CURRENT FILING DATE: 1998-08-03
;; EARLIER APPLICATION NUMBER: US 60/091,650
;; EARLIER FILING DATE: 1998-07-02
;; EARLIER APPLICATION NUMBER: US 60/054,646
;; EARLIER FILING DATE: 1997-08-04
;; NUMBER OF SEQ ID NOS: 18
;; SOFTWARE: FASTSEQ for Windows Version 3.0
;; SEQ ID NO 17
;; LENGTH: 176373
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)..(176373)
;; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Query Match 64.08; Score 16; DB 3; Length 176373;
Best Local Similarity 79.28; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 catcgccagattcgatcctcgt 25
||||| ||||| ||||| ||||| |||||
Db 97026 catcgccagattcattcttctgt 97049

RESULT 13
US-08-706-281A-17
;; Sequence 17, Application US/08706281A
;; Patent No. 6100048
;; GENERAL INFORMATION:
;; APPLICANT: Cone, Roger D
;; APPLICANT: Fan, Wei
;; APPLICANT: Boston, Bruce A
;; APPLICANT: Kesterton, Robert A
;; APPLICANT: Lu, Dongxi
;; APPLICANT: Chen, Mengbiao
;; TITLE OF INVENTION: Methods and Reagents for Discovering and
;; TITLE OF INVENTION: Using Mammalian Melanocortin Receptor Agonists and Antagonists
;; NUMBER OF SEQUENCES: 19
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
;; STREET: 300 South Wacker Drive
;; CITY: Chicago
;; STATE: IL
;; COUNTRY: USA
;; ZIP: 60606
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/706,281A
;; FILING DATE: 04-SEP-1996
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: No. 6100048nan, Kevin E
;; REGISTRATION NUMBER: 35,303
;; REFERENCE/DOCKET NUMBER: 96,886
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312-913-0001
;; TELEFAX: 312-913-0002
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 17:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 978 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear

;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..975
US-08-706-281A-17

Query Match 62.48; Score 15.6; DB 3; Length 978;
Best Local Similarity 81.88; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 catcgccagattcgatcctc 23
||||| ||||| ||||| ||||| |||||
Db 486 CATCGCCGATTCGATTCGACCTTC 507

RESULT 14
US-08-417-103-15
;; Sequence 15, Application US/08417103
;; Patent No. 5723299
;; GENERAL INFORMATION:
;; APPLICANT: Bell, Graeme I.
;; APPLICANT: Yamada, Yulchiro
;; APPLICANT: Seino, Susumu
;; TITLE OF INVENTION: SOMATOSTATIN RECEPTORS
;; NUMBER OF SEQUENCES: 16
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Arnold, White & Durkee
;; STREET: P.O. Box 4433
;; CITY: Houston
;; STATE: Texas
;; COUNTRY: United States of America
;; ZIP: 77210
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/417,103
;; FILING DATE: 05-APR-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/816,283
;; FILING DATE: 01-DEC-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Wilson, Mark B.
;; REGISTRATION NUMBER: 37,259
;; REFERENCE/DOCKET NUMBER: ARCD:144
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (512) 418-3000
;; TELEFAX: (512) 474-7577
;; INFORMATION FOR SEQ ID NO: 15:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1147 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 15..1121
US-08-417-103-15

Query Match 62.48; Score 15.6; DB 1; Length 1147;
Best Local Similarity 81.88; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 catcgccagattcgatcctc 23
||||| ||||| ||||| ||||| |||||
Db 242 CATCGCACATTTTACATCCTC 263

RESULT 15
US-07-816-283-5
Sequence 5, Application US/07816283
Patent No. 5436155
GENERAL INFORMATION:
APPLICANT: Bell, Graeme I.
APPLICANT: Yamada, Yuichiro
APPLICANT: Saino, Susumu
TITLE OF INVENTION: SOMATOSTATIN RECEPTORS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: PO Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/816,283
FILING DATE: 19911231
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: McDaniel, C. Steven
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1400
TELEFAX: 713-789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1351 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-07-816-283-5

Query Match 62.4%; Score 15.6; DB 1; Length 1351;
Best Local Similarity 81.8%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 catgccagcatttcgacccctc 23
Db 310 CATCACCAACATTTCACATCCTC 331

RESULT 16
US-08-417-103-5
Sequence 5, Application US/08417103
Patent No. 5723299
GENERAL INFORMATION:
APPLICANT: Bell, Graeme I.
APPLICANT: Yamada, Yuichiro
APPLICANT: Saino, Susumu
TITLE OF INVENTION: SOMATOSTATIN RECEPTORS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/417,103
FILING DATE: 05-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/816,283
FILING DATE: 01-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: ARCD:144
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1351 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 83..1189
US-08-417-103-5

Query Match 62.4%; Score 15.6; DB 1; Length 1351;
Best Local Similarity 81.8%; Pred. No. 1.2e+02;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 catgccagcatttcgacccctc 23
Db 310 CATCACCAACATTTCACATCCTC 331

RESULT 17
US-08-998-416-617
Sequence 617, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Reibischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPII
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwalls Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Weigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976

CITY: Houston
STATE: Texas
COUNTRY: US
ZIP: 77210
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,727A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/273,567
FILING DATE: 11-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:461/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
TELEX: N/A
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-479-727A-27

Query Match 61.6%; Score 15.4; DB 3; Length 1126;
Best Local Similarity 76.0%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcatcgccagcattcgatcctg 25
||| | ||||| | ||||| |
Db 796 GCACCTCCAGCATCTTGAGCTGT 772

RESULT 21
US-08-482-369A-27/c
Sequence 27, Application US/08482369A
Patent No. 6093399
GENERAL INFORMATION:
APPLICANT: Thorpe, Philip E.
APPLICANT: Edgington, Thomas S.
TITLE OF INVENTION: Methods and Compositions for the Specific
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: US
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,369A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/273,567
FILING DATE: 11-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.

REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD:433/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
TELEX: N/A
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 1126 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-482-369A-27

Query Match 61.6%; Score 15.4; DB 3; Length 1126;
Best Local Similarity 76.0%; Pred. No. 1.4e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcatcgccagcattcgatcctg 25
||| | ||||| | ||||| |
Db 796 GCACCTCCAGCATCTTGAGCTGT 772

RESULT 22
PCT-US95-07439-27/c
Sequence 27, Application PC/TUS9507439
GENERAL INFORMATION:
APPLICANT: BOARD OF REGENTS, THE UNIVERSITY OF
APPLICANT: TEXAS SYSTEM
STREET: 201 West 7th Street
CITY: Austin
APPLICANT: STATE: Texas
APPLICANT: COUNTRY: United States of America
APPLICANT: POSTAL CODE: 78701
APPLICANT: TELEPHONE NO: (512)499-4462
APPLICANT: TELEFAX: (512)499-4523
APPLICANT: NAME: THE SCRIPPS RESEARCH INSTITUTE
APPLICANT: STREET: 10666 North Torrey Pines Road
APPLICANT: CITY: LaJolla
APPLICANT: STATE: California
APPLICANT: COUNTRY: United States of America
APPLICANT: POSTAL CODE: 92037
TITLE OF INVENTION: METHODS AND COMPOSITIONS
FOR THE SPECIFIC
TITLE OF INVENTION: COAGULATION OF VASCULATURE
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: US
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS, ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07439
FILING DATE: Concurrently herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/273,567
FILING DATE: 11-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: PARKER, DAVID L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: UTSD433P--
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/762,308
FILING DATE: 09-DEC-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/224,593
FILING DATE: 05-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: UTSD.335--1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 418-3000
TELEFAX: 474-7577
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1956 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-762-308-10

Query Match 60.8%; Score 15.2; DB 2; Length 1956;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 gccatgccatgccatgccatc 20
||||| ||||| ||||| |||||
Db 1266 GCATCTCCAGCCTCTCGATC 1247

RESULT 26
US-08-299-849B-25/c
Sequence 25, Application US/08299849B
Patent No. 5612201
GENERAL INFORMATION:
APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
APPLICANT: Lech, Bernard; Szikora, Jean-Pierre; De Smet, Charles;
APPLICANT: Chomez, Patrick
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful in
TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/299,849B
FILING DATE: 1-SEPTEMBER-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/037,230
FILING DATE: 26-MARCH-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838

APPLICATION NUMBER: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 5612201man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5355
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 2099 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: smage-II
US-08-299-849B-25

Query Match 60.8%; Score 15.2; DB 1; Length 2099;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 6 gccatgccatgccatgccatc 25
||||| ||||| ||||| |||||
Db 944 CCACGATTTCACTCCTTGT 925

RESULT 27
US-08-142-368A-25/c
Sequence 25, Application US/08142368A
Patent No. 5925729
GENERAL INFORMATION:
APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
TITLE OF INVENTION: Rejection Antigens and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/142,368A
FILING DATE: 02-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991


```

:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 2099 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: genomic DNA
:   FEATURE:
:     NAME/KEY: smage-11
:   US-08-037-230D-25
:
: Query Match
:   Best Local Similarity 60.8%; Score 15.2; DB 4; Length 2099;
:   Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
:
: Oy 6 gccagcattcgatcctcgt 25
:     |||||
: Db 944 gccagcattcgatcctcgt 925
:
: RESULT 30
: US-08-299-849B-24/c
:   Sequence 24, Application US/08299849B
:   Patent No. 5612201
:   GENERAL INFORMATION:
:     APPLICANT: De Plaen, Etienne; Boon-Falleur, Thierry;
:     APPLICANT: Leth, Bernard; Szikora, Jean-Pierre; De Smet, Charles;
:     APPLICANT: Chomez, Patrick
:     TITLE OF INVENTION: Isolated Nucleic Acid Molecules Useful In
:     TITLE OF INVENTION: Determining Expression Of A Tumor Antigen Precursor
:     NUMBER OF SEQUENCES: 48
:     CORRESPONDENCE ADDRESS:
:       ADDRESSEE: Felfe & Lynch
:       STREET: 805 Third Avenue
:       CITY: New York City
:       STATE: New York
:       ZIP: 10022
:     COMPUTER READABLE FORM:
:       MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
:     COMPUTER: IBM
:     OPERATING SYSTEM: PC-DOS
:     SOFTWARE: Wordperfect
:     CURRENT APPLICATION DATA:
:       APPLICATION NUMBER: US/08/299,849B
:       FILING DATE: 1-SEPTEMBER-1994
:       CLASSIFICATION: 435
:       PRIOR APPLICATION DATA:
:         APPLICATION NUMBER: 08/037,230
:         FILING DATE: 26-MARCH-1993
:         PRIOR APPLICATION DATA:
:           APPLICATION NUMBER: PCT/US92/04354
:           FILING DATE: 22-MAY-1992
:           PRIOR APPLICATION DATA:
:             APPLICATION NUMBER: 07/807,043
:             FILING DATE: 12-DECEMBER-1991
:             PRIOR APPLICATION DATA:
:               APPLICATION NUMBER: 07/764,364
:               FILING DATE: 23-SEPTEMBER-1991
:               PRIOR APPLICATION DATA:
:                 APPLICATION NUMBER: 07/728,838
:                 APPLICATION NUMBER: 9-JULY-1991
:                 PRIOR APPLICATION DATA:
:                   APPLICATION NUMBER: 07/705,702
:                   FILING DATE: 23-MAY-1991
:                   ATTORNEY/AGENT INFORMATION:
:                     NAME: Hanson, No. 5612201man D.
:                     REGISTRATION NUMBER: 30,946
:                     REFERENCE/DOCKET NUMBER: LUD 5355
:                     TELECOMMUNICATION INFORMATION:
:                       TELEPHONE: (212) 688-9200
:                       TELEFAX: (212) 838-3884
:                     INFORMATION FOR SEQ ID NO: 24:
:                       SEQUENCE CHARACTERISTICS:

```

```

:
:   LENGTH: 2150 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: genomic DNA
:   FEATURE:
:     NAME/KEY: smage-1
:   US-08-299-849B-24
:
: Query Match
:   Best Local Similarity 85.0%; Score 15.2; DB 1; Length 2150;
:   Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
:
: Oy 6 gccagcattcgatcctcgt 25
:     |||||
: Db 933 gccagcattcgatcctcgt 914
:
: RESULT 31
: US-08-142-368A-24/c
:   Sequence 24, Application US/08142368A
:   Patent No. 5925729
:   GENERAL INFORMATION:
:     APPLICANT: Boon-Falleur, Thierry; Van der Bruggen, Thierry;
:     APPLICANT: Van den Eynde, Beno t; Van Pel, Aline; De Plaen, Etienne;
:     APPLICANT: Lurquin, Christophe; Chomez, Patrick; Traversari, Catia
:     TITLE OF INVENTION: Tumor Rejection Antigen Precursors, Tumor
:     TITLE OF INVENTION: Rejection Antigens and Uses Thereof
:     NUMBER OF SEQUENCES: 26
:     CORRESPONDENCE ADDRESS:
:       ADDRESSEE: Felfe & Lynch
:       STREET: 805 Third Avenue
:       CITY: New York City
:       STATE: New York
:       ZIP: 10022
:     COMPUTER READABLE FORM:
:       MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
:     COMPUTER: IBM
:     OPERATING SYSTEM: PC-DOS
:     SOFTWARE: Wordperfect
:     CURRENT APPLICATION DATA:
:       APPLICATION NUMBER: US/08/142,368A
:       FILING DATE: 02-MAY-1994
:       CLASSIFICATION: 435
:       PRIOR APPLICATION DATA:
:         APPLICATION NUMBER: PCT/US92/04354
:         FILING DATE: 22-MAY-1992
:         PRIOR APPLICATION DATA:
:           APPLICATION NUMBER: 07/807,043
:           FILING DATE: 12-DECEMBER-1991
:           PRIOR APPLICATION DATA:
:             APPLICATION NUMBER: 07/764,364
:             FILING DATE: 23-SEPTEMBER-1991
:             PRIOR APPLICATION DATA:
:               APPLICATION NUMBER: 07/728,838
:               APPLICATION NUMBER: 9-JULY-1991
:               PRIOR APPLICATION DATA:
:                 APPLICATION NUMBER: 07/705,702
:                 FILING DATE: 23-MAY-1991
:                 ATTORNEY/AGENT INFORMATION:
:                   NAME: Hanson, No. 5925729man D.
:                   REGISTRATION NUMBER: 30,946
:                   REFERENCE/DOCKET NUMBER: LUD 5253.4-US
:                   TELECOMMUNICATION INFORMATION:
:                     TELEPHONE: (212) 688-9200
:                     TELEFAX: (212) 838-3884
:                   INFORMATION FOR SEQ ID NO: 24:
:                     SEQUENCE CHARACTERISTICS:
:                       LENGTH: 2150 base pairs
:                       TYPE: nucleic acid
:                       STRANDEDNESS: single
:                       TOPOLOGY: linear

```

MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: smage-I
US-08-142-368A-24

Query Match 60.8%; Score 15.2; DB 2; Length 2150;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 gccagcattgcgcctcgt 25
|||||
DB 933 GCCAGCATTTCACTCCTTGT 914

RESULT 32
US-08-967-727-24/c
Sequence 24, Application US/08967727
Patent No. 6025474
GENERAL INFORMATION:
APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESS: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
CLASSIFICATION: 435
FILING DATE: 26-MARCH-1993
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,365
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6025474man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5553
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 2150 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: smage-I

US-08-967-727-24

Query Match 60.8%; Score 15.2; DB 3; Length 2150;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 gccagcattgcgcctcgt 25
|||||
DB 933 GCCAGCATTTCACTCCTTGT 914

RESULT 33
US-08-037-230D-24/c
Sequence 24, Application US/08037230D
Patent No. 6235525
GENERAL INFORMATION:
APPLICANT: Gaugler, B atrice; Van den Eynde, Beno t;
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Coding For
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESS: Felfe & Lynch
STREET: 805 Third Avenue
CITY: New York City
STATE: New York
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
CLASSIFICATION: 435
FILING DATE: 26-MARCH-1993
APPLICATION NUMBER: PCT/US92/04354
FILING DATE: 22-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/807,043
FILING DATE: 12-DECEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/764,364
FILING DATE: 23-SEPTEMBER-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/728,838
FILING DATE: 9-JULY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/705,702
FILING DATE: 23-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 6235525man D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5553
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 688-9200
TELEFAX: (212) 838-3884
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 2150 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
FEATURE:
NAME/KEY: smage-I
US-08-037-230D-24

Query Match 60.8%; Score 15.2; DB 4; Length 2150;
Best Local Similarity 85.0%; Pred. No. 1.9e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 6 gccagcattgcctcgt 25
|||||
Db 933 gccagcattgcctcgt 914

RESULT 34

US-08-693-214-1/c
; Sequence 1, Application US/08693214
; Patent No. 6054318

GENERAL INFORMATION:

APPLICANT: COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH
APPLICANT: ORGANISATION
TITLE OF INVENTION: Expression of the Glucose Oxidase Gene
TITLE OF INVENTION: in Transgenic Organisms
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOPER AND DUNHAM LLP
STREET: 1185 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/693,214
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PM 3792/94
FILING DATE: 10-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: INTERNATIONAL PCT/AU95/00059
FILING DATE: 10-FEB-1995
ATTORNEY/AGENT INFORMATION:
NAME: WHITE, JOHN P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 1 212 278 0400
TELEFAX: 1 212 391 0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 3029 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Talaromyces flavus
STRAIN: ATCC 32908

FEATURE:

NAME/KEY: CDS
LOCATION: 1032..2846
US-08-693-214-1

Query Match 60.8%; Score 15.2; DB 3; Length 3029;
Best Local Similarity 85.0%; Pred. No. 2e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 gccagcattgcctcgt 24
|||
Db 1305 gccagcattgcctcgt 1286

RESULT 35

US-08-404-665-3/c
; Sequence 3, Application US/08404665
; Patent No. 5591583

GENERAL INFORMATION:

APPLICANT: Reid, Robert A.
APPLICANT: Ackley, Rhonda L.
APPLICANT: Hemperly, John J.
TITLE OF INVENTION: HUMAN RESTRICTIN AND NUCLEIC ACID
TITLE OF INVENTION: SEQUENCES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
ADDRESSEE: Company
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ

COUNTRY: US

ZIP: 07417

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/404,665
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Fugit, Donna R.
REGISTRATION NUMBER: 32,135
REFERENCE/DOCKET NUMBER: P-3341
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4724 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-404-665-3

Query Match 60.8%; Score 15.2; DB 1; Length 4724;
Best Local Similarity 85.0%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 catcgagcattgcctcgt 21
|||
Db 903 catcgagcattgcctcgt 884

RESULT 36

US-08-404-671-3/c
; Sequence 3, Application US/08404671
; Patent No. 5635360

GENERAL INFORMATION:

APPLICANT: Reid, Robert A.
APPLICANT: Ackley, Rhonda L.
APPLICANT: Hemperly, John J.
TITLE OF INVENTION: HUMAN RESTRICTIN AND NUCLEIC ACID
TITLE OF INVENTION: SEQUENCES
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
ADDRESSEE: Company
STREET: 1 Becton Drive
CITY: Franklin Lakes
STATE: NJ

COUNTRY: US
ZIP: 07417

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/404,671

ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/247,475
FILING DATE: May 23, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..42
FEATURE:
NAME/KEY: CDS
LOCATION: 43..63
FEATURE:
NAME/KEY: CDS
LOCATION: 64..102
US-08-247-475-24

Query Match 60.0%; Score 15; DB 1; Length 102;
Best Local Similarity 78.3%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 catgcgcagcattcgatctcgcg 24
||||| ||||| ||||| ||
Db 24 CATGCCGCGCATCGGATCCTCG 46

RESULT 40
US-08-479-650-24
Sequence 24, Application US/08479650
Patent No. 5599544
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
TITLE OF INVENTION: Recombinant Infectious Bovine
TITLE OF INVENTION: Rhinotracheitis Virus
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/479,650
FILING DATE: June 7, 1995
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..42
FEATURE:
NAME/KEY: CDS
LOCATION: 43..63
FEATURE:
NAME/KEY: CDS
LOCATION: 64..102
US-08-479-650-24

Query Match 60.0%; Score 15; DB 1; Length 102;
Best Local Similarity 78.3%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 catgcgcagcattcgatctcgcg 24
||||| ||||| ||||| ||
Db 24 CATGCCGCGCATCGGATCCTCG 46

RESULT 41
US-08-191-866D-30
Sequence 30, Application US/08191866D
Patent No. 5783195
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
TITLE OF INVENTION: Recombinant Infectious Bovine
TITLE OF INVENTION: Rhinotracheitis Virus S-IBR-052 And Uses Thereof
NUMBER OF SEQUENCES: 99
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/191,866D
FILING DATE: 4 February 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 102 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..42
FEATURE:
NAME/KEY: CDS
LOCATION: 43..63
FEATURE:
NAME/KEY: CDS
LOCATION: 64..102
US-08-191-866D-30

Query Match 60.0%; Score 15; DB 1; Length 102;
Best Local Similarity 78.3%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 2 catgcgcagcattgcattcctcg 24
||||| ||| ||||| |
Db 24 CATGCCGCGCATCGGATCCTCG 46

RESULT 42
US-08-674-169-24
Sequence 24, Application US/08674169
Patent No. 5804372

GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: Macdonald, Richard D.
TITLE OF INVENTION: Recombinant Infectious Bovine
TITLE OF INVENTION: Rhinotracheitis Virus
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/674,169
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..42
FEATURE:

NAME/KEY: CDS
LOCATION: 43..63
FEATURE:
NAME/KEY: CDS
LOCATION: 64..102
US-08-674-169-24

Query Match 60.0%; Score 15; DB 1; Length 102;
Best Local Similarity 78.3%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 2 catgcgcagcattgcattcctcg 24
||||| ||| ||||| |
Db 24 CATGCCGCGCATCGGATCCTCG 46

RESULT 43
US-08-185-949B-30
Sequence 30, Application US/08185949B
Patent No. 5874279

GENERAL INFORMATION:
APPLICANT: Mark D. Cochran
APPLICANT: Richard D. Macdonald
TITLE OF INVENTION: Recombinant Infectious Bovine
TITLE OF INVENTION: Rhinotracheitis Virus
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM 330 466 DX2
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,949B
FILING DATE: 03-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 278-0525
INFORMATION FOR SEQ ID NO: 30:
SEQUENCE CHARACTERISTICS:
LENGTH: 102 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..42
FEATURE:
NAME/KEY: CDS
LOCATION: 43..63
FEATURE:
NAME/KEY: CDS
LOCATION: 64..102
US-08-185-949B-30

Query Match 60.0%; Score 15; DB 2; Length 102;
Best Local Similarity 78.3%; Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 2 catgccagcattgcattccg 24
||||| |||| |
Db 24 CATGCCGCGCATGCGATCTTGG 46

RESULT 44
US-08-328-322-3/c
Sequence 3, Application US/08328322
Patent No. 5723436
GENERAL INFORMATION:
APPLICANT: Huang, Laiqiang
APPLICANT: Cyert, Martha S.
TITLE OF INVENTION: Calcineurin Interacting Protein Compositions
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,322
FILING DATE: 24-OCT-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: P38,615
REFERENCE/DOCKET NUMBER: 8600-0151.10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 918 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: unknown
TOPOLOGY: unknown
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: CINC coding sequence
US-08-328-322-3

Query Match 60.0%; Score 15; DB 1; Length 918;
Best Local Similarity 78.3%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 3 atgcgcagcattgcattccgt 25
||||| |||| |
Db 408 ATGCCATCATGTGATCTTGGT 386

RESULT 45
US-08-328-322-1/c
Sequence 1, Application US/08328322
Patent No. 5723436
GENERAL INFORMATION:
APPLICANT: Huang, Laiqiang
APPLICANT: Cyert, Martha S.
TITLE OF INVENTION: Calcineurin Interacting Protein Compositions
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:

ADDRESS: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,322
FILING DATE: 24-OCT-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: P38,615
REFERENCE/DOCKET NUMBER: 8600-0151.10
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1222 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: Sau3AI fragment containing CNIC
FEATURE:
NAME/KEY: CDS
LOCATION: 1..918
US-08-328-322-1

Query Match 60.0%; Score 15; DB 1; Length 1222;
Best Local Similarity 78.3%; Pred. No. 2.2e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

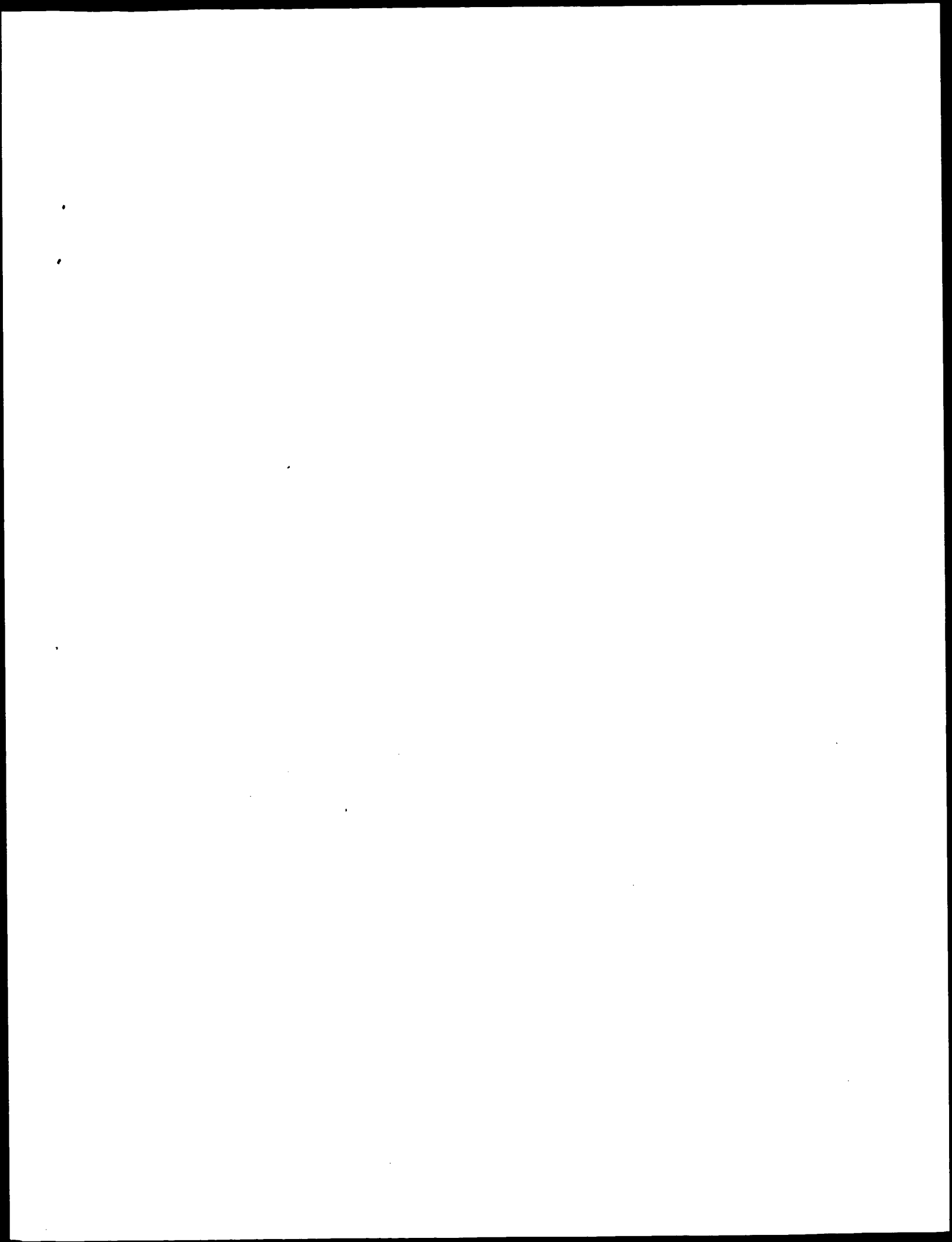
Oy 3 atgcgcagcattgcattccgt 25
||||| |||| |
Db 408 ATGCCATCATGTGATCTTGGT 386

Search completed: October 9, 2001, 11:39:57
Job time: 1912 sec

Wed Oct 10 07:44:49 2001

us-09-396-196f-7.std.rni

Page 21



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 13:46:43 ; Search time 5930.9 Seconds
(without alignments)
39.846 Million cell updates/sec

Title: US-09-396-196f-7

Perfect score: 25
Sequence: 1 gcacgcacgacatcgcacgcgcgt 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18.6	74.4	216	246	AZ577684	13b12 Sho
2	18.6	74.4	278	189	W10819	ma5h02.r1
3	18.6	74.4	538	30	AA439422	AA439422
4	18.6	74.4	636	113	AM202534	687038H06
5	18.6	73.6	617	118	AM637075	6154809..w
6	18.2	72.8	307	131	BB336604	BB336604
7	18.2	72.8	878	168	BF684643	BF684643
8	17.8	71.2	453	223	AQ068105	HS-2177..B
9	17.8	71.2	472	175	BG278235	a2f02np..f
10	17.8	71.2	568	21	AI519307	LD38845..5
11	17.8	71.2	606	14	AA941558	LD25751..5
12	17.8	71.2	616	250	AZ852429	2M0155A15
13	17.8	71.2	636	15	AI108881	GH08185..5
14	17.8	71.2	649	15	AI063994	GH04236..5
15	17.8	71.2	654	258	TA44D09P	AL455005.T..bruce1
16	17.6	70.4	208	113	AM203894	SMOV3MCM
17	17.6	70.4	295	5	AA335698	EST40140
18	17.6	70.4	353	112	AM172391	AM172391
19	17.6	70.4	360	14	AA994865	o0b6e08..s
20	17.6	70.4	400	1	AA067753	mm56a05..r
21	17.6	70.4	442	4	AA294431	SMOV3MCM
22	17.6	70.4	448	234	AQ068181	n0eb0026J
23	17.6	70.4	519	166	BE372749	601224261
24	17.6	70.4	529	112	AM182566	x144b11..x
25	17.6	70.4	541	222	FR000291	Z86694.F..ruidpries
26	17.6	70.4	554	114	AM288131	SMOV3MCM
27	17.6	70.4	706	112	AM184046	AM184046
28	17.6	70.4	790	106	AL578914	F018a10..Y
29	17.6	70.4	879	145	BF161425	601770406
30	17.6	70.4	920	106	AL556716	AL556716
31	17.6	70.4	1022	151	BF608598	MY1.00154
32	17.4	69.6	281	104	AI978168	6140A1AF1
33	17.4	69.6	299	115	AM355901	707015F11
34	17.4	69.6	385	188	T14661	05c04a07-f2
35	17.4	69.6	395	102	AI834428	606068A02
36	17.4	69.6	471	102	AI812153	946059F01
37	17.4	69.6	504	136	BE511349	606011H09
38	17.4	69.6	525	23	AI692089	606011H09
39	17.4	69.6	528	24	AI740036	605070A01
40	17.4	69.6	536	24	AI714408	606013D04
41	17.4	69.6	550	102	AI834659	606071E07
42	17.4	69.6	555	102	AI861139	603012P12
43	17.4	69.6	565	118	AM600443	660068C06
44	17.4	69.6	573	102	AI833970	605097G09
45	17.4	69.6	582	23	AI68324	605023G02

ALIGNMENTS

RESULT 1
 LOCUS AZ577684 216 bp DNA GSS 08-DEC-2000
 DEFINITION 13b12 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
 sp. NGR234 genomic clone 13b12, DNA sequence.

ACCESSION AZ577684
 VERSION AZ577684.1 GI:11604206
 KEYWORDS GSS.
 SOURCE Rhizobium sp. NGR234.
 ORGANISM Rhizobium sp. NGR234.
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 Rhizobiaceae; Rhizobium.

REFERENCE 1 (bases 1 to 216)
 AUTHORS Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.
 TITLE Genetic snapshots of the Rhizobium species NGR234 genome
 JOURNAL Genomics.com 1 (6), 0014.1-0014.7 (2000)
 COMMENT Contact: Virginie Viprey

Laboratoire de Biologie Molculaire des Plantes Supérieures
 University of Geneva
 1 Chemin de l'Imperatrice, Chambes/Geneva 1292, Switzerland
 Tel: +44(0)1603450000
 Fax: +44(0)1603450045
 Email: virginie.viprey@bbsrc.ac.uk
 Class: shotgun.

FEATURES

source location/Qualifiers
 1..216
 /organism="Rhizobium sp. NGR234"
 /strain="ANU265"
 /db_xref="taxon:394"
 /clone="13b12"
 /clone_lib="Shot-gun genomic library of Rhizobium strain
 ANU265"
 /note="Vector: M13; derivative strain of NGR234 cured of
 pNGR234a"
 BASE COUNT 34 a 70 c 65 g 46 t 1 others
 ORIGIN

Query Match 74.4% Score 18.6; DB 246; Length 216;
 Best Local Similarity 84.0% Pred. No. 1.1e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcatcgccagatccgacccgt 25
 Db 114 GCATCGCCAGCATGCGACACCAT 138

RESULT 2

LOCUS W10819/c
 DEFINITION ma5h02.r1 Soares mouse p3NMf19.5 Mus musculus cDNA clone
 IMAGE:314547.5 similar to gb:J04982.cdsl ADP,ATP CARRIER PROTEIN,
 HEART/SKELETAL MUSCLE ISOFORM (HUMAN);, mRNA sequence.

ACCESSION W10819
 VERSION W10819.1 GI:1285135

KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.

REFERENCE 1 (bases 1 to 278)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.

TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT Contact: Maria M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of MedicineP
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LML; contact the
 IMAB Consortium (info@lml.gov) for further information.
 MGI: 205163
 Trace considered overall poor quality
 Seq primer: mob,REGA+ET
 High quality sequence stop: 1.

FEATURES

source location/Qualifiers
 1..278
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:314547"
 /clone_lib="Soares mouse p3NMf19.5"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="PH10B (ampicillin resistant)"
 /note="Vector: pT73D (Pharmacia) with a modified

QY	1	gcattgccagcatttcgattccgtcgt 25	74.4%; Score 18.6; DB 189; Length 278;
Db	198	gctgcacacgacatttcgattccgtcgt 174	Best Local Similarity 84.0%; Pred. No. 1.1e+02; Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
RESULT	3	AVA394322 538 bp mRNA EST 23-AUG-2000	
LOCUS	AVA394322/c		
DEFINITION	AVA394322	Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone	
ACCESSION	PS050c10_r 5'	mRNA sequence.	
VERSION	AVA394322		
KEYWORDS	AVA394322.1 GI:8594647		
SOURCE	EST.		
ORGANISM	Porphyra yezoensis.		
REFERENCE	Porphyra yezoensis.		
AUTHORS	Eukariyola; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;		
TITLE	Porphyra.		
JOURNAL	1 (bases 1 to 538)		
MEDLINE	Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and		
COMMENT	Tabata,S.		
	Generation of 10,154 expressed sequence tags from a leafy		
	gametophyte of a marine red alga, Porphyra yezoensis		
	20363100		
	Contact: Erika Asamizu		
	The First Laboratory for Plant Gene Research		
	Kazusa DNA Research Institute		
	Yana 1532-3, Kisarazu, Chiba 292-0812, Japan		
	Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.		
FEATURES	Location/Qualifiers		
SOURCE	1..538		
	/organism="Porphyra yezoensis"		
	/strain="TU-1"		
	/db_xref="taxon:2788"		
	/clone="PS050c10_r"		
	/clone_lib="Porphyra yezoensis TU-1"		
	/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:		
	XhoI"		
BASE COUNT	114 a 158 c 170 g 96 t		
ORIGIN			
Query Match			
Best Local Similarity	84.0%;		
Matches	21; Conservative	0; Mismatches	4; Indels 0; Gaps 0;
QY	1	gcattgccagcatttcgattccgtcgt 25	
Db	198	gctgcacacgacatttcgattccgtcgt 174	
RESULT	3	AVA394322 538 bp mRNA EST 23-AUG-2000	
LOCUS	AVA394322/c		
DEFINITION	AVA394322	Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone	
ACCESSION	PS050c10_r 5'	mRNA sequence.	
VERSION	AVA394322		
KEYWORDS	AVA394322.1 GI:8594647		
SOURCE	EST.		
ORGANISM	Porphyra yezoensis.		
REFERENCE	Porphyra yezoensis.		
AUTHORS	Eukariyola; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;		
TITLE	Porphyra.		
JOURNAL	1 (bases 1 to 538)		
MEDLINE	Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and		
COMMENT	Tabata,S.		
	Generation of 10,154 expressed sequence tags from a leafy		
	gametophyte of a marine red alga, Porphyra yezoensis		
	20363100		
	Contact: Erika Asamizu		
	The First Laboratory for Plant Gene Research		
	Kazusa DNA Research Institute		
	Yana 1532-3, Kisarazu, Chiba 292-0812, Japan		
	Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.		
FEATURES	Location/Qualifiers		
SOURCE	1..538		
	/organism="Porphyra yezoensis"		
	/strain="TU-1"		
	/db_xref="taxon:2788"		
	/clone="PS050c10_r"		
	/clone_lib="Porphyra yezoensis TU-1"		
	/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:		
	XhoI"		
BASE COUNT	114 a 158 c 170 g 96 t		
ORIGIN			
Query Match			
Best Local Similarity	84.0%;		
Matches	21; Conservative	0; Mismatches	4; Indels 0; Gaps 0;
QY	1	gcattgccagcatttcgattccgtcgt 25	
Db	198	gctgcacacgacatttcgattccgtcgt 174	
RESULT	3	AVA394322 538 bp mRNA EST 23-AUG-2000	
LOCUS	AVA394322/c		
DEFINITION	AVA394322	Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone	
ACCESSION	PS050c10_r 5'	mRNA sequence.	
VERSION	AVA394322		
KEYWORDS	AVA394322.1 GI:8594647		
SOURCE	EST.		
ORGANISM	Porphyra yezoensis.		
REFERENCE	Porphyra yezoensis.		
AUTHORS	Eukariyola; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;		
TITLE	Porphyra.		
JOURNAL	1 (bases 1 to 538)		
MEDLINE	Nikaido,I., Asamizu,E., Nakajima,M., Nakamura,Y., Saga,N. and		
COMMENT	Tabata,S.		
	Generation of 10,154 expressed sequence tags from a leafy		
	gametophyte of a marine red alga, Porphyra yezoensis		
	20363100		
	Contact: Erika Asamizu		

ACCESSION	sequence.
VERSION	AM202534
KEYWORDS	AM202534.1 GI:6501161
SOURCE	EST.
ORGANISM	Zea mays.
	Zea mays.
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
REFERENCE	Clade: Panicoideae; Andropogoneae; Zea.
AUTHORS	1 (bases 1 to 636)
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL	Unpublished (1999)
COMMENT	Contact: Walbot V Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 725 8221 Email: walbot@stanford.edu Plate: 687038 row: H column: 06. Location/Qualifiers 1..636 /organism="Zea mays" /cultivar="Illinois High Oil" /db_xref="taxon:4577" /clone_lib="687 - Early embryo from Delaware" /rsluse_type="embryo" /dev_stage="14, 21, 28, and 35 days after pollination" /lab_host="E. coli SOLR" /note="Organ: embryo; Vector: pBluescript SK; Site: 1; Xho: the Uni-ZAP XR system (Stratagene BN937328-12). Clones were picked by a Q-bot after blue/white selection (ampicillin resistance - use 100 micograms/microliter). developed from a pool of equal amounts of RNA from developing embryos sampled at 14, 21, 28 and 35 days after pollination of the Illinois High Oil Maize strain Cycle 90. This closed strain has been selected for high oil concentration for 90 generations and originates from the 1890s era open pollinated variety Burr's White"
BASE COUNT	111 a 197 c 190 g 138 t
ORIGIN	
Query Match	74.4%; Score 18.6; DB 113; Length 636;
Best Local Similarity	84.0%; Pred. No. 1.2e+02;
Matches 21; Conservative	0; Mismatches 4; Indels 0; Gaps 0;
QY	1 gcatgcagcagcatctgcactcgt 25
DB	578 GCAGCGCCAGCAATGCAGCCYGT 554
RESULT 5	
LOCUS	AM637075/c
DEFINITION	AM637075 617 bp mRNA EST 03-APR-2000
ACCESSION	bl54a09.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
VERSION	laevis cDNA clone PBX0054A09 5', mRNA sequence.
KEYWORDS	AM637075.1 GI:7394175
SOURCE	EST.
ORGANISM	African clawed frog.
	Xenopus laevis
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
	Xenopodinae; Xenopus.
REFERENCE	1 (bases 1 to 617)
AUTHORS	Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Stafla,N.G., Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M., Touchman J.W., Bonaldo,M.F. and Soares,M.B.
TITLE	THE XENUS Maternal EST project

QY 1 gcatgcagcatcgcctc 23
 Db 56 GCATGCTAGCATTCGATCTC 34

RESULT 7
 LOCUS BF684643 878 bp mRNA EST 22-DEC-2000
 DEFINITION 602141001F1 NIH-MGC-46 Homo sapiens cDNA clone IMAGE:4301856 5',
 mRNA sequence.
 ACCESSION BF684643
 VERSION BF684643
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: ggapds-rt@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L16M1163 row: b column: 01
 High quality sequence stop: 689.

FEATURES
 source
 1. 878
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4301856"
 /clone_11b="NIH-MGC-46"
 /tissue="type="leiomyosarcoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pOTB7; Site:1: XhoI; Site:2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Size-selected >500bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-CDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH-MGC
 Library."

BASE COUNT 250 a 202 c 306 g 120 t
 ORIGIN

Query Match 72.8%; Score 18.2; DB 168; Length 878;
 Best Local Similarity 87.0%; Pred. No. 2e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 atccgagcatcgcctcgt 25
 Db 788 ATCGTCAGCCTTCGATCCACGT 766

RESULT 8
 LOCUS AQ068105 453 bp DNA GSS 04-AUG-1998
 DEFINITION HS.2177.B2.E02.MR.C1T Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate-2177 Col-4 Row-J, DNA sequence.
 ACCESSION AQ068105
 VERSION AQ068105.1 GI:3379373
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS

TITLE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 JOURNAL 1 (bases 1 to 453)
 MEDLINE Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 COMMENT Koller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589

Contact: Mahairas Gg, Wallace Jc, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 2177 row: J column: 4
 Class: BAC ends
 High quality sequence stop: 453.

FEATURES
 source

1. 453
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-2177 Col-4 Row-J"
 /clone_11b="C1T Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBel0BAC11; BAC clones in
 E-Coli DH10B"
 BASE COUNT 117 a 108 c 73 g 152 t 3 others
 ORIGIN

Query Match 71.2%; Score 17.8; DB 223; Length 453;
 Best Local Similarity 90.5%; Pred. No. 2.8e+02;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 tcgccagcatcgcctcgt 24
 Db 418 TCGCAGCAGCTTCGATCTCG 438

RESULT 9
 LOCUS BG278235 472 bp mRNA EST 21-FEB-2001
 DEFINITION a2102np.r1 Neurospora crassa sexual cDNA library, uni-zap vector
 system Neurospora crassa cDNA clone a2102np 5', mRNA sequence.
 ACCESSION BG278235
 VERSION BG278235.1 GI:13074338
 KEYWORDS EST.
 SOURCE Neurospora crassa.
 ORGANISM Neurospora crassa.

REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariales; Sordariaceae; Neurospora.
 AUTHORS Kupfer,D., Lai,H., Nelson,M. and Roe,B.
 TITLE ESTs from a Neurospora crassa Sexual cDNA Library
 JOURNAL Unpublished (2001)
 COMMENT Other-ESTs: a2102np.f1
 Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu

Contact Dr. Mary Anne Nelson, Department of Biology, University of
 New Mexico, Albuquerque, NM 87131 (e-mail address manelson@unm.edu)
 regarding clone availability
 Seq primer: M13 Universal Reverse Primer
 High quality sequence stop: 421.
 Location/Qualifiers

```

//color_cis="TADACH":7.22/
//clone="LD38845"
//sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="Xtl Blue"
/note="Organ: embryo; Vector: pOT2; Site_1: EcoRI; Site_2: XhoI. Sized fractionated cDNAs were directly ligated into pOT2."
BASE COUNT      168 a   145 c    191 g     64 t
ORIGIN

```

Query Match	71.2%;	Score 17.8;	DB 21;	Length 568;
Best Local Similarity	90.5%;	Pred. No. 2.9e+02;		
Matches 19; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Email: http://www.frutifly.org/EST_est@frutifly.berkeley.edu
hit genomic sequence DS02777.4_e1
Plate: 257 row: E column: 3
High quality sequence spot: 549.

FEATURES

Source

1. .606
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD25751"
/sequence.lib="LD Drosophila melanogaster embryo port2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="X11 Blue"
/note="Jorgan: embryo; Vector: port2; Site_1: EcoRI; Site_2
XhoI; Sized fractionated cDNAs were directly ligated into
port2."

BASE COUNT

165 a 170 c 195 g 76 t

BRIGIN

		71.2%	Score 17.8:	DB 14:	Length 606;
Query Match		Best Local Similarity	90.5%:	Pred. No. 2.	9e+02:
Matches	19;	Conservative	0;	Mismatches	2;
				Indels	0;
Gy	1 gcatgcagcatttcgac	21			
Db	517 GCATGCCAGCATTCGAC	497			

RESULT	12
AZ852429	
LOCUS	AZ852429 616 bp DNA 21-FEB-2001
DEFINITION	CmouseU55A15F Mouse 10kb plasmid UUGC1M library Mus musculus genomic clone UUGC2M0155A15 F, DNA sequence.
ACCESSION	AZ852429
VERSION	AZ852429.1 GI:13039507
KEYWORDS	GSS.
SOURCE	house mouse.
ORGANISM	Mus musculus.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus 1 (bases 1 to 616)
AUTHORS	Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Rellly,M., Rose,M., Rose,R., Stokes,K., Tinney,A., von Niederhausern,A.

/note="Organ: head; Vector: POT2; Site.1: EcoRI; Site.2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library."

BASE COUNT 186 a 167 c 191 g 105 t

ORIGIN

Query Match 71.2%; Score 17.8; DB 15; Length 649;
Best Local Similarity 90.5%; Pred. No. 3e+02; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gcatcgccagcatctgcattcc 21
|||||
Db 530 GCATCGCCAGCATCTGCATCC 510

RESULT 15

LOCUS TAA4D09P 654 bp DNA GSS 13-DEC-2000
DEFINITION T. brucei sheared genomic DNA clone 44D09, forward sequence,
genomic survey sequence.

ACCESSION AL455005.1 GI:11856458

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 654)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.

TITLE

Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk

COMMENT

Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TRE0927/4 G0rat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaubin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

source

1..654
/organism="Trypanosoma brucei"
/strain="TRE0927"
/db_xref="taxon:5691"
/clone="44D09"

BASE COUNT 157 a 167 c 152 g 177 t 1 others

ORIGIN

Query Match 71.2%; Score 17.8; DB 258; Length 654;
Best Local Similarity 90.5%; Pred. No. 3e+02; 2; Indels 0; Gaps 0;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 gcatcgccagcatctgcattcc 21
|||||
Db 431 GCATCGCCAGCAATTCATCC 411

RESULT 16

LOCUS

DEFINITION AM203894 208 bp mRNA EST 02-DEC-1999
SMOV3MCA32A09SK Onchocerca volvulus molting L3 larva cDNA
(SL96MIM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCA32A09 5',

mRNA sequence.

ACCESSION AM203894.1 GI:6503412

VERSION

KEYWORDS

SOURCE

ORGANISM

Onchocerca volvulus.
Onchocerca volvulus.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.

1 (bases 1 to 208)

REFERENCE Williams, S.A., Lizotte-Waniewski, M., Laney, S. and Lustigman, S.
Genes expressed in molting L3 larvae of Onchocerca volvulus
Unpublished (1997)

TITLE

JOURNAL

COMMENT

Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 413/853/826
Fax: 413/853/786
Email: genome@smith.edu

Seq primer: pbluescript SK.
Location/Qualifiers

FEATURES

source

1..208

/organism="Onchocerca volvulus"

/strain="Kumba, Cameroon"

/db_xref="taxon:6282"

/clone="SMOV3MCA32A09"

/dev_stage="molting L3"

(SL96MIM-Ovml3)"

/lab_host="XLI-Blue MRF"

/note="Vector: lambda Uni-ZAP XR; Site.1: Eco RI; Site.2:

Xho I; Filarial nematode parasite of humans. Third-stage

larvae, L3, were isolated from infected black flies in

Cameroun (forest strain). The L3 were cultured in 20% FCS

in IMEM+ NCTC 135 and collected after day 1, 2, or 3 in

culture. L3 of O. volvulus molt to fourth-stage larvae by

day 5 in culture. mRNA was isolated from approximately

6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3

in culture, and converted to double-stranded cDNA using

reverse transcriptase and oligo(dT) followed by RNase H

and DNA pol I. The library was constructed in the lambda

Uni-Zap XR vector and has 1 x 10⁶ independent

recombinants and the average insert size is ~1200 bp. The

library was constructed by Sara Lustigman and Michelle

Lizotte-Waniewski in the laboratory of Dr. S. A. Williams.

The library is available from Dr. Sara Lustigman (email:

slustig@nybc.org).

BASE COUNT 76 a 40 c 57 g 35 t

ORIGIN

Query Match 70.4%; Score 17.6; DB 113; Length 208;
Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcatcgccagcatctgcattcc 24
|||||
Db 56 GCATCGCCACCTCCTCGATCCTCG 33

RESULT 17

LOCUS

DEFINITION AA335698 295 bp mRNA EST 21-APR-1997
EST40140 Epididymus Homo sapiens cDNA 5' end, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

1 (bases 1 to 295)
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS
 Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bull, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Well, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fitchman, J.L., Geoghegan, N.S., Glodde, A., Gnanapavan, C., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palacios, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Wiley, J., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferris, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

TITLE
 Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence

JOURNAL
 Nature 377 (6547 Suppl), 3-174 (1995)

COMMENT
 Other ESTs: THC181529
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES
 Location/Qualifiers
 1..295
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):137373"
 /clone_xref="taxon:9606"
 /clone_lib="Epididymus"
 /sex="male"
 /dev_stage="adult"
 /note="Organ: epididymis; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
 ECORI: Site_2: XhoI

BASE COUNT
 62 a 97 c 79 g 54 t 3 others

ORIGIN

Query Match 70.4%; Score 17.6; DB 5; Length 295;
 Best Local Similarity 83.3%; Pred. No. 3.3e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 2 catcgccagcatttcgacccctg 25
 ||||||| | |||||
 150 CATGCCAGCTTCTTCTATCTCTCT 173

RESULT 18
 AM172391 353 bp mRNA EST 16-NOV-1999
 LOCUS
 DEFINITION
 IMAGE:2659481.3 similar to TR:Q12799 Q12799 T-COMPLEX PROTEIN 10A
 ; mRNA sequence.
 AM172391
 AM172391.1 GI:6438339
 EST.

ACCESSION
 AM172391
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 353)
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40UP from Gibco
 High quality sequence stop: 345.

FEATURES
 Location/Qualifiers
 1..353
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2659481"
 /clone_lib="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1 M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Palma Bonaldo."

BASE COUNT
 66 a 88 c 106 g 93 t

ORIGIN

Query Match 70.4%; Score 17.6; DB 112; Length 353;
 Best Local Similarity 83.3%; Pred. No. 3.4e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 1 gcatcgccagcatttcgacccctg 24
 ||||||| | |||||
 210 GCATGCCAGCTTCTTGACCCCTCG 187

RESULT 19
 AA994865 360 bp mRNA EST 27-AUG-1998
 LOCUS
 DEFINITION
 IMAGE:162510.3' similar to TR:Q12799 Q12799 T-COMPLEX PROTEIN 10A.
 ; contains element THR repetitive element ; mRNA sequence.
 AA994865
 AA994865.1 GI:3181354
 EST.

ACCESSION
 AA994865
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 360)
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 637 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 305.
 Location/Qualifiers
 1..360
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:162510"
 /clone_lib="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI;

Equal amounts of plasmid DNA from three normalized libraries (fetal lung MBL19W, testis NHT, and B-cell NCI-CGAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 60 a 95 c 117 g 88 t

Query Match 70.4% Score 17.6; DB 14; Length 360;
Best Local Similarity 83.3%; Pred. No. 3.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcatcgccagcatctgcctcg 24
||||||| | | | | | | |
Db 261 GCATCGCCAGCTTCTGCACCTCG 238

RESULT 20
AA067753 400 bp mRNA EST 06-FEB-1997
LOCUS mms5a05.r1 Striatogene mouse embryonic carcinoma (#937317) Mus
DEFINITION musculus cDNA clone IMAGE:532400 5', mRNA sequence.
ACCESSION AA067753
VERSION AA067753.1 GI:1566033
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 400)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellingberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theisinger, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE The WashU-HMMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project
WashU-HMMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LIND; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:319336
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 389.

FEATURES

Source
1. 400
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:532400"
/clone_lib="Striatogene mouse embryonic carcinoma (#937317)"
/tissue_type="carcinoma"
/dev_stage="embryonic"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. P19 cell
line. Average insert size: 1.0 kb; uni-ZAP XR Vector; ~5'
adaptor sequence: 5' GAATTCGCGACGAC 3' ~3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTT 3' ~3' adaptor
72 a 123 c 125 g 80 t

BASE COUNT
ORIGIN

Query Match 70.4% Score 17.6; DB 1; Length 400;
Best Local Similarity 83.3%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcatcgccagcatctgcctcg 24
||||||| | | | | | | |
Db 219 GCATCGCCAGCGCTCCCTCTCG 196

RESULT 21
AA294431 442 bp mRNA EST 12-NOV-1997
LOCUS SMOV3CAB69SK Onchocerca volvulus molting L3 larva cDNA
DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3CAB69 5', mRNA
sequence.
ACCESSION AA294431
VERSION AA294431.1 GI:2099620
KEYWORDS EST.
SOURCE Onchocerca volvulus.
ORGANISM Onchocerca volvulus

REFERENCE
AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
1 (bases 1 to 442)
Williams, S.A., Lizotte-Maniowski, M., Laney, S. and Lustigman, S.
Genes expressed in molting L3 larvae of Onchocerca volvulus
Unpublished (1997)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.

FEATURES
Source
1. 442
Location/Qualifiers

/organism="Onchocerca volvulus"
/strain="Kumba, Cameroons"
/db_xref="taxon:6282"
/clone="SMV3CAB69"
/clone_lib="Onchocerca volvulus molting L3 larva cDNA
(SL96MLM-Ovml3)"
/dev_stage="molting L3"
/lab_host="XLI-Blue MR"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
Xho I; Filarial nematode parasite of humans. Third-stage
larvae, L3, were isolated from infected black flies in
Cameroon (forest strain). The L3 were cultured in 20% FCS
in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
culture. L3 of O. volvulus molt to fourth-stage larvae by
day 5 in culture. mRNA was isolated from approximately
6000 molting larvae (ml3), 2000 larvae from day 1, 2 or 3
in culture, and converted to double-stranded cDNA using
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library was constructed in the lambda
Uni-Zap XR vector and has 1 x 10⁶ independent
recombinants and the average insert size is ~1200 bp. The
library was constructed by Sara Lustigman and Michelle
Lizotte-Maniowski in the laboratory of Dr. S. A. Williams.
The library is available from Dr. Sara Lustigman (email:
slustig@mc.biology.edu)."

BASE COUNT 138 a 80 c 123 g 86 t 15 others

ORIGIN

Query Match 70.4% Score 17.6; DB 4; Length 442;
Best Local Similarity 83.3%; Pred. No. 3.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY 1 gcatcgccagcatctgcctcg 24
||||||| | | | | | | |

DB 50 GCATGCCACCTCTCGATCTCG 27

RESULT 22
 A0868181/C
 LOCUS nbe00026j13r CUGI Rice BAC Library (EcoRI) Oryza sativa genomic
 DEFINITION clone nbe00026j13r, DNA sequence.
 ACCESSION A0868181
 VERSION A0868181.1 GI:6218736
 KEYWORDS GSS.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Euphorbiales; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Oryza.
 REFERENCE 1 (bases 1 to 488)
 AUTHORS Wing, R.A. and Dean, R.A.
 TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
 JOURNAL Unpublished (1998)
 COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: GGAAACACCTATGACCATG
 Class: BAC ends
 High quality sequence start: 32
 High quality sequence stop: 328.
 Location/Qualifiers
 1..488
 /organism="Oryza sativa"
 /strain="Japanica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="nbe00026j13r"
 /clone_lib="CUGI Rice BAC Library (EcoRI)"
 /tissue_type="leaf"
 /lab_host="E. coli DH10B"
 /note="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
 Rice is the most important food crop in the world. Half of
 the world population, especially those inhabiting highly
 populated areas of the humid tropics and subtropics, rely
 on rice as their primary source of carbohydrate.
 Monocotyledonous rice is a diploid plant (2n=24) with a
 haploid genome equivalent of 431 Mbp (Arumuganathan and
 Earle, 1991). The relatively small genome of rice, three
 times larger than that of Arabidopsis, makes it suitable
 for genomic studies. In order to facilitate positional
 cloning, physical mapping and genome sequencing of rice,
 we have constructed a BAC library from Oryza sativa,
 Nipponbare variety using EcoRI as the cloning enzyme. The
 library contains 55,296 clones with an average insert size
 of 121 kb providing approximately 15 haploid genome
 equivalents. The deep coverage allows the isolation a
 particular sequence with a probability of 99.9%. Three
 high density filters, each containing 18,432 clones
 (doubly spotted), represent the whole library for colony
 screening and can be requested from the Clemson University
 BAC/EST Resource Center (www.genome.clemson.edu)."

FEATURES

SOURCE

BASE COUNT 86 a 163 c 143 g 94 t 2 others
 ORIGIN

Query Match 70.4%; Score 17.6; DB 234; Length 488;
 Best Local Similarity 83.3%; Pred. No. 3.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 catcgccagcattcgatctcg 25
 ||||| ||||| ||||| ||||| ||
 DB 104 CATCTCCAGCGCTTCCATCTTGT 81

RESULT 23

BE372749/c
 LOCUS BE372749
 DEFINITION BE372749
 ACCESSION BE372749
 VERSION BE372749.1 GI:9318112
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 519)
 AUTHORS NIH-MGC
 TITLE NIH-MGC
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
 COMMENT Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM8738 row: 1 column: 15
 High quality sequence stop: 519.
 Location/Qualifiers
 1..519
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:3582806"
 /clone_lib="NCI CGAP Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="3 months, virgin"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

FEATURES

SOURCE

BASE COUNT 104 a 145 c 158 g 110 t 2 others
 ORIGIN

Query Match 70.4%; Score 17.6; DB 166; Length 519;
 Best Local Similarity 83.3%; Pred. No. 3.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcatcgccagcattcgatctcg 24
 ||||| ||||| ||||| ||||| ||
 DB 206 GCATGCCACGCGCTCTCG 183

RESULT 24

AM182566/c
 LOCUS AM182566
 DEFINITION xj44b11.x1 Soares.NFL.T GBC.S1 Homo sapiens cDNA clone
 IMAGE:2660061 3' similar to TR:012799 G12799 T-COMPLEX PROTEIN 10A
 ; contains element M81 repetitive element ;, mRNA sequence.
 ACCESSION AM182566
 VERSION AM182566.1 GI:6451026
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 529)
 AUTHORS NCI-CGAP
 TITLE NCI-CGAP
 Tumor Gene Index

JOURNAL COMMENT

Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -400P from Gibco
 High quality sequence stop: 463.

FEATURES

Location/Qualifiers
 1..529

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2660061"
 /clone_lib="Soares_NFL_T_GBC_S1"
 /lab_host="DH10B"
 /note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBH119W, testis NHT, and B-cell NCI-GAP-GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of 1.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT
 100 a 118 c 176 g 135 t
 ORIGIN

Query Match
 Best Local Similarity 83.3%; Score 17.6; DB 112; Length 529;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcacgcacgacatcgcacccg 24
 ||||||||| || |||||
 Db 260 GCATGCCAGCTTCTGACCCCTCG 237

RESULT 25
 FR0002911 541 bp DNA GSS 27-FEB-1997
 LOCUS F.rubripes GSS sequence, clone 015A17bF6, genomic survey sequence.
 DEFINITION Z86694
 VERSION Z86694.1 GI:1883606
 KEYWORDS GSS; genome survey sequence.
 SOURCE Takifugu rubripes.
 ORGANISM Takifugu rubripes.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Tetraodontiformes; Tetraodontidae; Takifugu.

REFERENCE
 1 (bases 1 to 541)
 Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrantia,Y., Williams,G. and Brenner,S.
 TITLE Direct Submission
 JOURNAL Submitted (18-FEB-1997) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridgeshire, CB10 1SB. Email: b10ne1@hgmrc.mrc.ac.uk
 COMMENT Vector: pBluescript II KS
 V-type: phagemid
 PRIMER: KS
 DESCR: One pass dye-terminator sequencing of cosmid cloned genomic sequence.

FEATURES

Location/Qualifiers
 1..541

/organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone_lib="cosmid 015A17"
 /clone="015A17bF6"
 BASE COUNT
 120 a 168 c 95 g 142 t 16 others
 ORIGIN

Query Match
 Best Local Similarity 70.4%; Score 17.6; DB 222; Length 541;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 catcgccagcatcgcacccg 25
 || |||| ||||| |||||
 Db 297 CAGCGCCCTCATTTCCACCTCGT 320

RESULT 26
 AW288131 554 bp mRNA EST 16-JAN-2000
 LOCUS SMOV3MCA34B09SK Onchocerca volvulus molting L3 larva cDNA
 DEFINITION (SL96MLM-Ovml3) Onchocerca volvulus cDNA clone SMOV3MCA34B09 5', mRNA sequence.
 ACCESSION
 VERSION AW288131 GI:6695066
 KEYWORDS EST.
 SOURCE Onchocerca volvulus.
 ORGANISM Onchocerca volvulus.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidae; Onchocercidae; Onchocerca.

REFERENCE
 1 (bases 1 to 554)
 Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
 TITLE Genes expressed in molting L3 larvae of Onchocerca volvulus
 JOURNAL Unpublished (1997)
 COMMENT Contact: Steven A. Williams
 Molecular Parasitology
 Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: pBluescript SK.

FEATURES

Location/Qualifiers
 1..554

/organism="Onchocerca volvulus"
 /strain="Kumba, Cameroons"
 /db_xref="taxon:6282"
 /clone="SMOV3MCA34B09"
 /clone_lib="Onchocerca volvulus molting L3 larva cDNA (SL96MLM-Ovml3)"
 /dev_stage="molting L3"
 /lab_host="X11-Blue MRF"
 /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. Third-stage larvae, L3, were isolated from infected black flies in Cameroon (forest strain). The L3 were cultured in 20% FCS in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in culture. L3 of O. volvulus molt to fourth-stage larvae by day 5 in culture. mRNA was isolated from approximately 6000 molting larvae (mJ3), 2000 larvae from day 1, 2 or 3 in culture, and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library was constructed in the lambda Uni-ZAP XR vector and has 1 x 10⁶ independent recombinants and the average insert size is ~1200 bp. The library was constructed by Sara Lustigman and Michelle Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams. The library is available from Dr. Sara Lustigman (email: slustigmen@bc.org)."
 BASE COUNT
 182 a 99 c 142 g 131 t
 ORIGIN

Query Match
 Best Local Similarity 70.4%; Score 17.6; DB 114; Length 554;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 OY 1 gcacgcacgacatcgcacccg 24
 ||||||||| | |||||||||

DB 55 GCATGCCATCTCTCATCTCTG 32

RESULT 27
AM184046/c
LOCUS
DEFINITION AM184046 706 bp mRNA EST 18-NOV-1999
f08a10.y1 zebrafish adult olfactory Danilo rerio cdna 5' similar to
tr-095071 095071 PROGESTIN INDUCED PROTEIN.; mRNA sequence.

ACCESSION AM184046
VERSION AM184046.1 GI:6452560
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danilo rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Rasbora; Danilo.

REFERENCE 1 (bases 1 to 706)
Clark, M., Johnson, S.L., Lebrach, H., Lee, R., Li, F., Marra, M., Eddy,
S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood,
K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rilter, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.,
and Wilson, R.
WashU zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
CDNA Library Preparation: John Ngai CDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
Resourcezentrum Primatendatenbank, Berlin, Germany (web address:
www.rzp.dz)
Seq primer: T3 ET from Amersham
High quality sequence stop: 489.

FEATURES
Source
1. 706
/organism="Danilo rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish adult olfactory"
/sex="mixed"
/tissue_type="olfactory rosettes"
/dev_stage="adult"
/lab_host="DI0Hb (Gibco BRL)"
/note="Vector: pSPORT1; Site 1: NotI; Site 2: SalI; This
is a directionally cloned cdna library from adult
zebrafish olfactory epithelium."

BASE COUNT 152 a 204 c 211 g 139 t

ORIGIN

Query Match 70.4%; Score 17.6; DB 112; Length 706;
Best Local Similarity 83.3%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

DB 329 CATGCCATCTCTCATCTCTG 306

RESULT 28
AL578914/c
LOCUS
DEFINITION AL578914 790 bp mRNA EST 16-FEB-2001
prime, mRNA sequence.
ACCESSION AL578914
VERSION AL578914.1 GI:12943445

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 790)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cdna libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segreg@genoscope.cns.fr; Web : www.genoscope.cns.fr.

FEATURES
Source
1. 790
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="GSDK005YE06"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="Placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand CDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded CDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : liang@lifetech.com URL :
<http://fulllength.invitrogen.com>"

BASE COUNT 195 a 212 c 234 g 132 t 17 others

ORIGIN

Query Match 70.4%; Score 17.6; DB 106; Length 790;
Best Local Similarity 83.3%; Pred. No. 3.8e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

DB 650 CATGCCATCTCTCATCTCTG 627

RESULT 29
BF161425/c
LOCUS
DEFINITION BF161425 879 bp mRNA EST 30-OCT-2000
601770406F1 NCI_CGAP_Lu29 Mus musculus cdna clone IMAGE:3989880 5',
mRNA sequence.
ACCESSION BF161425
VERSION BF161425.1 GI:11041578
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 879)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
Plate: L1AM9200 row: k column: 01
High quality sequence stop: 726.

FEATURES
Source
1. 879
/organism="Mus musculus"

/strain="CZECH II (fetal)"
 /db_xref="taxon:10090"
 /clone_image="3989880"
 /clone_lib="NCI CGAP_Lu29"
 /tissue_type="spontaneous tumor, metastatic to mammary."
 Stem cell origin."
 /lab_host="DH10B"
 /note="Organ: Lung; Vector: PCMV-SPORT6; Site_1: Salt;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 191 a 238 c 310 g 140 t

ORIGIN

Query Match 70.4%; Score 17.6; DB 145; Length 879;
 Best Local Similarity 83.3%; Pred. No. 3.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 catgccagcattcgatccgct 25
 ||| ||||| ||||| ||| |
 Db 424 CATGCCAGCTTTCGATGCTCCT 401

RESULT 30
 LOCUS AL556716 920 bp mRNA EST 16-FEB-2001
 DEFINITION AL556716 L7L_NFL006.PL2 Homo sapiens CDNA clone CSODK005YE06 5
 prime, mRNA sequence.
 ACCESSION AL556716
 VERSION AL556716.1 GI:12899649
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 920)
 TITLE Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 JOURNAL Unpublished (2001)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers

FEATURES
 source
 1..920
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CSODK005YE06"
 /clone_lib="L7L_NFL006.PL2"
 /tissue_type="placenta"
 /note="Vector: PCMVSPORT 6; Site_1: NotI; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the PCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"

BASE COUNT 203 a 320 c 224 g 172 t 1 others

ORIGIN

Query Match 70.4%; Score 17.6; DB 106; Length 920;
 Best Local Similarity 83.3%; Pred. No. 3.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 catgccagcattcgatccgct 25
 ||||| ||| | ||||| ||| |
 Db 769 CATGCCAGCTTTCGATGCTCCT 792

RESULT 31
 LOCUS BF608598/c 1022 bp mRNA EST 01-APR-2001
 DEFINITION MY_001540 Mouse 9-day fetus cDNA library ICRFP522 Mus musculus
 cDNA clone ICRFP522P0150 5', mRNA sequence.
 ACCESSION BF608598
 VERSION BF608598.1 GI:13505090
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1022)
 TITLE Yanyan, M., Hennig, S., Neidhardt, L., Radelof, U., Hermann, B.G.,
 Lehrach, H. and O'Brien, J.
 JOURNAL Detection of a high number of novel genes in a 9-day mouse embryo
 cDNA library normalised by oligonucleotide fingerprinting
 COMMENT Unpublished (2001)
 Contact: Hennig S
 laboratory 123, dept. Lehrach
 Max-Planck-Institut fuer Molekulare Genetik
 Ihmestr.65-73, D-14195 Berlin, Germany
 Tel: +49 30 8413 1612
 Fax: +49 30 8413 1380
 Email: hennig@molgen.mpg.de
 EST's are made from clones being representatives of clone clusters.
 Clone clusters were calculated from oligonucleotide fingerprints.
 PCR Primers
 FORWARD: 5'-GAGCTATTCCAGAGTACTGA-3'
 BACKWARD: 5'-TAATACGACTCCTACTATAGG-3'
 Seq primer: 5'-ATTAGTGACACTATAG-3'
 High quality sequence stop: 1022.
 Location/Qualifiers

FEATURES
 source
 1..1022
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="ICRFP522P0150"
 /clone_lib="Mouse 9-day fetus cDNA library ICRFP522"
 /tissue_type="whole embryo"
 /dev_stage="embryonic 9-day"
 /lab_host="E.coli, Xli blue"
 /note="Vector: PSVSPORT1; Site_1: NotI; Site_2: SalI;
 library preparation by oligo-dT priming of RNA. Clones can
 be ordered from the Resource Center in Berlin,
 http://www.rzpd.de."

BASE COUNT 229 a 298 c 268 g 227 t

ORIGIN

Query Match 70.4%; Score 17.6; DB 151; Length 1022;
 Best Local Similarity 83.3%; Pred. No. 3.9e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 gcatgccagcattcgatccgct 24
 ||||| ||||| ||| |||||
 Db 250 GCATGCCAGCTTCGCTCCTCG 227

RESULT 32
 LOCUS A1978168/c 281 bp mRNA EST 27-AUG-1999
 DEFINITION 614041A04.x2 614 - root cDNA library from Waldbot Lab Zea mays cDNA,
 mRNA sequence.
 ACCESSION A1978168
 VERSION A1978168.1 GI:5791376
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 281)

AUTHORS
Walbot, V.
Maize ESTs from various cDNA libraries sequenced at Stanford University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 614041 row: A column: 04.
Location/Qualifiers

FEATURES
Source
1..281
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="614 - root cDNA library from Walbot lab"
/tissue_type="root"
/dev_stage="3-4 days old"
/lab_host="X10LR"
/note="Organ: root; Vector: pBluescriptII SK+; Site: 1; EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot Lab (LM)"
Lab (LM)"

BASE COUNT
74 a 79 c 67 g 60 t 1 others

Query Match 69.6%; Score 17.4; DB 104; Length 281;
Best Local Similarity 94.7%; Pred. No. 4.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 107 CATGCCCTGCATTTCGATC 89
||||| |||||||

RESULT 33
AM355901 299 bp mRNA EST 03-FEB-2000
LOCUS 707015F11.x1 707 - Mixed adult tissues from Walbot lab (SK) Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION AM355901
VERSION AM355901.1 GI:6859907
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 299)
Walbot, V.
Maize ESTs from various cDNA libraries sequenced at Stanford University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 707015 row: F column: 11.
Location/Qualifiers

FEATURES
Source
1..299
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="707 - Mixed adult tissues from Walbot lab (SK)
/tissue_type="tassel, kernel, silk, husk, root, leaf"
/dev_stage="adult"
/lab_host="DH10B"

/note="Organ: tassel, kernel, silk, husk, root, leaf;
Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
differentiated maize tissues from an active Mutator
plant. Tissue ratio is 4/2/1/1/1 (tassel, kernel, silk,
husk, root, leaf). Unidirectionally cloned."
husk, root, leaf)"

BASE COUNT
71 a 74 c 84 g 68 t 2 others

Query Match 69.6%; Score 17.4; DB 115; Length 299;
Best Local Similarity 94.7%; Pred. No. 4.2e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 214 CATGCCCTGCATTTCGATC 232
||||| |||||||

RESULT 34
T14661/c 385 bp mRNA EST 17-OCT-1996
LOCUS 05c04a07-f21 membrane-free polysomes from endosperm Zea mays cDNA
DEFINITION clone 05c04a07 3' end similar to sucrose synthase, mRNA sequence.
ACCESSION T14661
VERSION T14661.1 GI:440640
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 385)
Shen, B., Carneiro, N., Torres-Jerez, I., Stevenson, R., McCreery, T.,
Helentjaris, T., Baysdorfer, C., Almira, E., Ferl, R., Hadden, J. and
Larkins, B.
Partial sequencing and mapping of clones from two maize cDNA
libraries
Plant Mol. Biol. 26, 1085-1101 (1994)
95111093
Other ESTs: 05c04a07-t7
Contact: The Maize CDNA Project

Helentjaris TG (primary contact)
Dept. of Plant Sciences
University of Arizona
Dept. of Plant Sciences, University of Arizona, Tucson, AZ, 85721
ph: 602-6218-746
fax: 602-621-7186
E-mail: helentjars@cvt.arizona.edu

Chris Baysdorfer
Department of Biological Sciences, School of Science
Hayward, CA 94542
ph: 510-881-3459
fax: 510-727-2035
E-mail: cbaysdor@esl.csu Hayward.edu

Rob Ferl
Interdisciplinary Center for Biotechnology Research
DNA Sequencing Core
University of Florida
P. O. Box 100695
Gainesville, FL 32611-0695
ph: 904-392-1928, ext. 301
fax: 904-392-4072
E-mail: robferl@nervm.nerdc.ufl.edu
Seq primer: F21.
Location/Qualifiers

FEATURES
Source
1..385
/organism="Zea mays"
/strain="W64A2"
/db_xref="taxon:4577"

```

/clone="05c04a07"
/clone_lib="membrane-free polysomes from endosperm"
/lab_host="DH10B"
/notes="Vector: Ziplox; Site_1: SalI; Site_2: NotI; ds-cDNA
was prepared from oligo-dT selected mRNA by priming with a
NotI oligo - dT oligomer and then adding the second strand
to RNase-nicked DNA:RNA hybrid with DNA Pol. SalI
adaptors were added to the ends, the ds-cDNAs were then
digested with NotI and size-selected. These were
directionally-cloned into the ziplox phage vector, excised
as plasmids, and then individually analyzed."

BASE COUNT      115 a      107 c      88 g      75 t
ORIGIN

Query Match      69.6%; Score 17.4; DB 188; Length 385;
Best Local Similarity 94.7%; Pred. No. 4.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2 catgccagcattcgatc 20
        ||||||| |||||||||
Db      184 CATGCCCTGCATTTCGATC 166

RESULT 35
A1834428      395 bp      mRNA      EST      02-FEB-2000
LOCUS      606068A02.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
DEFINITION      mays cDNA, mRNA sequence.
ACCESSION      A1834428
VERSION      A1834428.1 GI:5468637
KEYWORDS      EST.
SOURCE      Zea mays.
ORGANISM      Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 395)
AUTHORS      Walbot V.
TITLE      Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL      Unpublished (1999)
COMMENT      Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 606068 row: A column: 02.

FEATURES
    source
        1..395
        /organism="Zea mays"
        /cultivar="Ohio43"
        /db_xref="taxon:4577"
        /clone_lib="606 - Ear tissue cDNA library from Schmidt
        lab"
        /rissue_type="mixed"
        /dev_stage="ear length from 0.5 cm - 2.0 cm"
        /lab_host="XLOLR (Stratagene)"
        /note="Organ: immature ear; Vector: PBK-CMV; Site_1: EcoRI
        ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
        lab"

BASE COUNT      115 a      99 c      94 g      87 t
ORIGIN

Query Match      69.6%; Score 17.4; DB 102; Length 395;
Best Local Similarity 94.7%; Pred. No. 4.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2 catgccagcattcgatc 20
        ||||||| |||||||||

```

```

Db      229 CATGCCCTGCATTTCGATC 211

RESULT 36
A1812153      471 bp      mRNA      EST      02-FEB-2000
LOCUS      605087H09.y1 605 - Endosperm cDNA library from Schmidt lab Zea mays
DEFINITION      cDNA, mRNA sequence.
ACCESSION      A1812153
VERSION      A1812153.1 GI:5398809
KEYWORDS      EST.
SOURCE      Zea mays.
ORGANISM      Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 471)
AUTHORS      Walbot V.
TITLE      Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL      Unpublished (1999)
COMMENT      Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605087 row: H column: 09.

FEATURES
    source
        1..471
        /organism="Zea mays"
        /cultivar="Ohio43"
        /db_xref="taxon:4577"
        /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
        /rissue_type="nucellar, embryo, and endosperm"
        /dev_stage="10-14 days post-pollination"
        /lab_host="DH5(alpha)"
        /note="Organ: Kernel; Vector: pMD-GAL4-2'; Site_1: EcoRI;
        Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
        lab"

BASE COUNT      124 a      106 c      116 g      125 t
ORIGIN

Query Match      69.6%; Score 17.4; DB 102; Length 471;
Best Local Similarity 94.7%; Pred. No. 4.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      2 catgccagcattcgatc 20
        ||||||| |||||||||
Db      203 CATGCCCTGCATTTCGATC 221

RESULT 37
BE511349      504 bp      mRNA      EST      07-AUG-2000
LOCUS      946059F01.y1 946 - tassal primordium prepared by Schmidt lab Zea
DEFINITION      mays cDNA, mRNA sequence.
ACCESSION      BE511349
VERSION      BE511349.1 GI:9732597
KEYWORDS      EST.
SOURCE      Zea mays.
ORGANISM      Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 504)
AUTHORS      Walbot V.
TITLE      Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL      Unpublished (1999)
COMMENT      Contact: Walbot V

```

Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946059 row: F column: 01.
Location/Qualifiers

FEATURES

source

1. .504
/organism="Zea mays"
/cultivar="OH43"
/db.xref="taxon:4577"
/clone_lib="946 - tassels primordium prepared by Schmidt lab"
/tissue_type="tassels"
/dev_stage="just after the transition from vegetative to inflorescence development"
/lab_host="XLOLR"

/note="Organ: tassels; Vector: HybridAP; Site_1: EcoRI; Site_2: XhoI; George Chuck dissected immature tassels between 1mm and 3mm. Sharon Stanfield prepared the cDNA library in HybridAP. Sample insert size range was 350 bp to 3 kb with a 1 kb average."

BASE COUNT

122 a 113 c 133 g 136 t

ORIGIN

Query Match 69.6%; Score 17.4; DB 136; Length 504;
Best Local Similarity 94.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 catgccagcattgcgac 20
||||| |||||||
Db 276 CATGCCCTGCATTTCGATC 294

RESULT 38
A1692089/c 525 bp mRNA EST 02-FEB-2000
LOCUS 606011H09.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION A1692089
VERSION A1692089.1 GI:4967187
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 525)
Walbot.V.
Maize ESTs from various cDNA libraries sequenced at Stanford University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 606011 row: H column: 09.
Location/Qualifiers

JOURNAL

Unpublished (1999)

Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 606011 row: H column: 09.
Location/Qualifiers

FEATURES

source

1. .525
/organism="Zea mays"
/cultivar="Ohio43"
/db.xref="taxon:4577"
/clone_lib="606 - Ear tissue cDNA library from Schmidt lab"
/tissue_type="mixed"
/dev_stage="ear length from 0.5 cm - 2.0 cm"
/lab_host="XLOLR (Stratagene)"
/note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI

/note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI

BASE COUNT 142 a 137 c 124 g 121 t 1 others
ORIGIN ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt lab"

Query Match 69.6%; Score 17.4; DB 23; Length 525;
Best Local Similarity 94.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 catgccagcattgcgac 20
||||| |||||||
Db 231 CATGCCCTGCATTTCGATC 213

RESULT 39
A1740036 528 bp mRNA EST 02-FEB-2000
LOCUS 605070A01.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
DEFINITION cDNA, mRNA sequence.
ACCESSION A1740036
VERSION A1740036.1 GI:5108323
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 528)
Walbot.V.
Maize ESTs from various cDNA libraries sequenced at Stanford University
Unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605070 row: A column: 01.
Location/Qualifiers

REFERENCE 1 (bases 1 to 528)
AUTHORS Walbot V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
CONTACT: Walbot V
DEPARTMENT OF BIOLOGICAL SCIENCES
STANFORD UNIVERSITY
855 CALIFORNIA AVE, PALO ALTO, CA 94304, USA
TEL: 650 723 2227
FAX: 650 725 8221
EMAIL: walbot@stanford.edu
PLATE: 605070 row: A column: 01.
LOCATION/QUALIFIERS

FEATURES

source

1. .528
/organism="Zea mays"
/cultivar="Ohio43"
/db.xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/note="Organ: Kernel; Vector: PAD-GAL4-2'; Site_1: EcoRI; Site_2: XhoI; Kernel endosperm cDNA library from Schmidt lab"
BASE COUNT 137 a 148 c 133 g 109 t 1 others
ORIGIN

Query Match 69.6%; Score 17.4; DB 24; Length 528;
Best Local Similarity 94.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 catgccagcattgcgac 20
||||| |||||||
Db 144 CATGCCCTGCATTTCGATC 126

RESULT 40
A1714408 536 bp mRNA EST 02-FEB-2000
LOCUS 606013D04.x2 606 - Ear tissue cDNA library from Schmidt lab Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION A1714408
VERSION A1714408.1 GI:5018215

KEYWORDS EST
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 536)
Maize ESTs from various cDNA libraries sequenced at Stanford
TITLE University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 606013 row: D column: 04.

FEATURES
source
1..536
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="606 - Ear tissue cDNA library from Schmidt
lab"
/rissue_type="mixed"
/dev_stage="ear length from 0.5 cm - 2.0 cm"
/lab_host="XLOLR (Stratagene)"
/note="Organ: immature ear; Vector: pBK-CMV; Site:1: EORI
; Site:2: XhoI; Mixed ear tissue cDNA library from Schmidt
lab"

BASE COUNT 144 a 142 c 125 g 124 t 1 others
ORIGIN

Query Match 69.6%; Score 17.4; DB 24; Length 536;
Best Local Similarity 94.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 catgccagcattcgatc 20
||||| |||||||
DB 239 CATGCCCTGCATTTCGATC 221

RESULT 41
AI834659 550 bp mRNA EST 02-FEB-2000
LOCUS 606071E07.xl 606 - Ear tissue cDNA library from Schmidt lab Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION AI834659
VERSION AI834659.1 GI:5468868
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 550)
Maize ESTs from various cDNA libraries sequenced at Stanford
TITLE University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 606071 row: E column: 07.
Location/Qualifiers
1..550

FEATURES
source

/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="606 - Ear tissue cDNA library from Schmidt
lab"
/rissue_type="mixed"
/dev_stage="ear length from 0.5 cm - 2.0 cm"
/lab_host="XLOLR (Stratagene)"
/note="Organ: immature ear; Vector: pBK-CMV; Site:1: EORI
; Site:2: XhoI; Mixed ear tissue cDNA library from Schmidt
lab"

BASE COUNT 145 a 143 c 132 g 130 t
ORIGIN

Query Match 69.6%; Score 17.4; DB 102; Length 550;
Best Local Similarity 94.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 catgccagcattcgatc 20
||||| |||||||
DB 232 CATGCCCTGCATTTCGATC 214

RESULT 42
AI861139 555 bp mRNA EST 19-JUL-1999
LOCUS 603012F12.xl 603 - stressed root cDNA library from Wang/Bohnert lab
DEFINITION Zea mays cDNA, mRNA sequence.
ACCESSION AI861139
VERSION AI861139.1 GI:5525300
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 555)
Maize ESTs from various cDNA libraries sequenced at Stanford
TITLE University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 603012 row: F column: 12.
Location/Qualifiers
1..555
/organism="Zea mays"
/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="603 - stressed root cDNA library from
Wang/Bohnert lab"
/rissue_type="seedling"
/dev_stage="salt stress"
/lab_host="E. coli XL Gold"
/note="Organ: root; Vector: pBluescriptII SK(+); XR;
Seedling stressed root cDNA library from Wang/Bohnert lab"

BASE COUNT 140 a 148 c 137 g 130 t
ORIGIN

Query Match 69.6%; Score 17.4; DB 102; Length 555;
Best Local Similarity 94.7%; Pred. No. 4.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 catgccagcattcgatc 20
||||| |||||||
DB 214 CATGCCCTGCATTTCGATC 196

RESULT 43
LOCUS AM600443 565 bp mRNA EST 23-MAR-2000
DEFINITION 660068C06.xl 660 - Mixed stages of anther and pollen Zea mays cDNA,
mRNA sequence.
ACCESSION AM600443
VERSION AM600443.1 GI:7304504
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 565)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 660068 row: C column: 06.

FEATURES
source
1..565
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="660 - Mixed stages of anther and pollen"
/tissue_type="whole premeiotic anthers to pollen shed"
/dev_stage="premeiotic anthers to pollen shed"
/lab_host="XLOLR"
/note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
Site_2: XhoI; Anther and pollen cDNA library.
Directionally sequenced with 5' end at the EcoRI site.
Created by Amie Franklin."

BASE COUNT 148 a 161 c 141 g 115 t
ORIGIN

Query Match 69.6%; Score 17.4; DB 118; Length 565;
Best local Similarity 94.7%; Pred. No. 4.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 catgccagcattcgatc 20
||||| |||||||
Db 167 CATGCCCTGCATTTCGATC 149

RESULT 44
LOCUS AI833970 573 bp mRNA EST 02-FEB-2000
DEFINITION 605097G09.xl 605 - Endosperm cDNA library from Schmidt lab Zea mays
cDNA, mRNA sequence.
ACCESSION AI833970
VERSION AI833970.1 GI:5456280
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 573)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V

Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605097 row: G column: 09.

FEATURES
source
1..573
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/note="Organ: Kernel; Vector: PAD-GAL4-2'; Site_1: EcoRI;
Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

BASE COUNT 153 a 153 c 144 g 123 t
ORIGIN

Query Match 69.6%; Score 17.4; DB 102; Length 573;
Best local Similarity 94.7%; Pred. No. 4.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 catgccagcattcgatc 20
||||| |||||||
Db 225 CATGCCCTGCATTTCGATC 207

RESULT 45
LOCUS AI668324 582 bp mRNA EST 02-FEB-2000
DEFINITION 605023G02.xl 605 - Endosperm cDNA library from Schmidt lab Zea mays
cDNA, mRNA sequence.
ACCESSION AI668324
VERSION AI668324.1 GI:4827632
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 582)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605023 row: G column: 02.

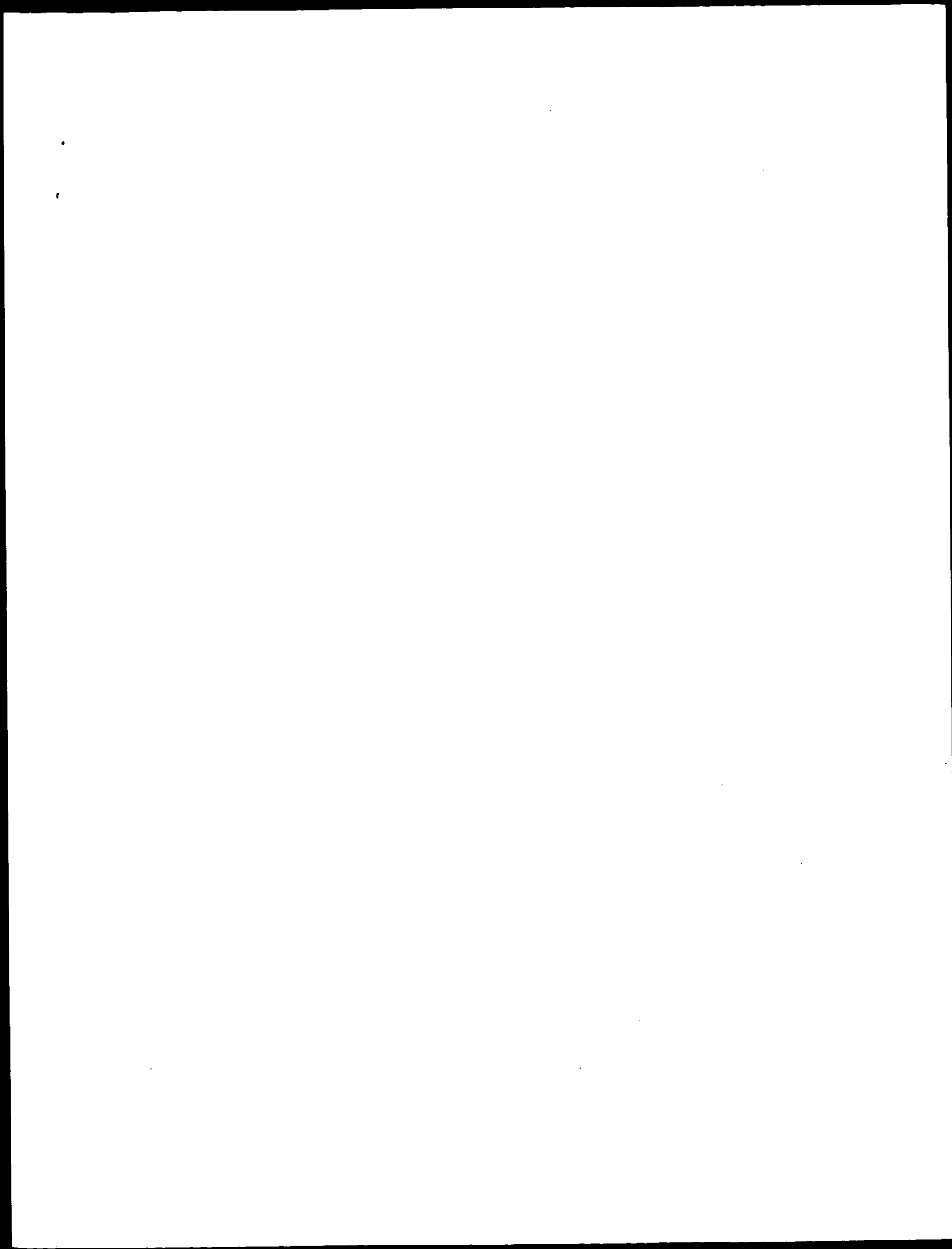
FEATURES
source
1..582
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/note="Organ: Kernel; Vector: PAD-GAL4-2'; Site_1: EcoRI;
Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

BASE COUNT 154 a 165 c 147 g 116 t
ORIGIN

Query Match 69.6%; Score 17.4; DB 23; Length 582;
 Best Local Similarity 94.7%; Pred. No. 4.6e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 catgcagcagcattcgatc 20
 ||||| ||||| ||||| |||||
 Db 182 CATGCCCTGCATTTGCATC 164

Search completed: October 9, 2001, 13:46:46
 Job time: 9521 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 12:13:27 : Search time 1666.31 Seconds
(without alignments)
232.066 Million cell updates/sec

Title: us-09-396-196f-8

Perfect score: 25
Sequence: 1 gcattcgatcctcgtcaggtgcag 25

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_hrgo_hum:*
20: em_hrgo_inv:*
21: em_hrgo_rtd:*
22: em_hrgo_hum1:*
23: em_hrg_hum2:*
24: em_hrg_hum3:*
25: em_hrg_hum4:*
26: em_hrg_hum5:*
27: em_hrg_hum6:*
28: em_hrg_hum7:*
29: em_hrg_hum8:*
30: em_hrg_inv1:*
31: em_hrg_inv2:*
32: em_hrg_other:*
33: em_hrg_rtd:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vl:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vl1:*
59: gb_vl2:*
60: gb_hrg1:*
61: gb_hrg2:*
62: gb_hrg3:*
63: gb_hrg4:*
64: gb_hrg5:*
65: gb_hrg6:*
66: gb_hrg7:*
67: gb_hrg8:*
68: gb_hrg9:*
69: gb_hrg10:*
70: gb_hrg11:*
71: gb_hrg12:*
72: gb_hrg13:*
73: gb_hrg14:*
74: gb_hrg15:*
75: gb_hrg16:*
76: gb_hrg17:*
77: gb_hrg18:*
78: gb_hrg19:*
79: gb_hrg20:*
80: gb_hrg21:*
81: gb_hrg22:*
82: gb_hrg23:*
83: gb_hrg24:*
84: gb_hrg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_pr10:*
95: gb_pr11:*
96: gb_pr12:*
97: gb_pr13:*
98: em_ba3:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	25	100.0	AR029499	AR029499 Sequence
2	25	100.0	AR034916	AR034916 Sequence
3	25	100.0	A11530	A11530 B10B gene o
4	25	100.0	E00893	E00893 Genomic DNA
5	25	100.0	ECOBIO	J04423 E.coli 7.8-
6	25	100.0	A38246	A38246 Sequence 1
7	25	100.0	A38251	A38251 Sequence 6
8	25	100.0	A93674	A93674 Sequence 1

Db 122 GCATTTCGATCCTCGTCGACGTCGAC 146

RESULT 4
LOCUS E00893 1121 bp DNA PAT 29-SEP-1997
DEFINITION Genomic DNA encoding biotin Synthetase.
ACCESSION E00893
VERSION E00893.1 GI:2169154
KEYWORDS JP 1986149091-A/1.
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1121)
AUTHORS Hirono,Y., Kojima,T. and Kimura,H.
TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN
JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
NIPPON SODA CO LTD
COMMENT OS Escherichia coli
PN JP 1986149091-A/1
PD 07-JUL-1986
PF 24-DEC-1984 JP 1984272605
PI HIROMO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1:19),(C12P13/18, PC
C12R1:19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=Escherichia coli Nsl01;
CC Feature is identified by experimental;
FH Key Location/Qualifiers
FT CDS 42..1079
/product=biotin synthetase'.
FEATURES
source 1..1121
/organism="Escherichia coli"
/db_xref="taxon:562"
BASE COUNT 289 a 296 c 325 g 211 t
ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 1121;
Best local Similarity 100.0%; Pred. No. 0.066;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaattcgatccctcgatcaggtcgac 25
|||||
Db 140 GCATTTCGATCCTCGTCGACGTCGAC 164

RESULT 5
LOCUS E00810 5793 bp DNA BCT 28-FEB-1994
DEFINITION E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase (biob), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc protein, and dehydrobiotin synthetase (biocD), complete cds.
ACCESSION J04423.1 GI:145422
VERSION J04423
KEYWORDS 7,8-diamino-pelargonic acid aminotransferase;
7-keto-8-amino-pelargonic acid synthetase; bioA gene; biob gene; bioc gene; biocD gene; bioF gene; biotin synthetase; dehydrobiotin synthetase.
SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5793)
AUTHORS Otsuka,A.J., Buoncrisiani,M.R., Howard,P.K., Flamm,J. and Johnson,O.

TITLE The Escherichia coli biotin biosynthetic enzyme sequences predicted
JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
MEDLINE 89066784
COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.
FEATURES
source 1..5793
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
complement(98..574)
/note="ORF 1"
/codon_start=1
/transl_table=1
/protein_id="AAA23513.1"
/db_xref="GI:455168"
/translation="MKLISNDLRDGLKPLRHVNGMGYDGDNI SPHLAMDVPAGTK SFVATGCDPADPTGSGWMMWVNLDPADRVLPFGSGSLVAMPDGLQTRDFKGTG YDGAAPKGETGTHRYIFTVHALDIERIDVDEGASAAAMGCVNHFHSLASITAMFS"
complement(633..1925)
/gene="bioA"
complement(633..1925)
/gene="bioA"
/codon_start=1
/transl_table=1
/product="7,8-diamino-pelargonic acid aminotransferase"
/protein_id="AAA23514.1"
/db_xref="GI:457106"
/translation="MTYDILAFDQRHIMPTYSMTSPLPVYVSAEGCELLISDGR ALVDGSMMAAIIHGYNHPOLNAAMKSO IDAMSHVMGCIHAPAI ELCRLVAMSGRN ALVDFLADSGSAVAEAMKALQYQAKGEARQPLTFENGYGDPFGMSVCDND SMHSIMKGYLPENLFAPQSRMDGDEMDNGFALMAHHEIAVAIIEPIVGA GGMHMYPEMLKTRKRCDBREGITLIDATATGRTGKLFACHEAIEADILCKKA LTGGTMTLSATLTITREVAERTISNGEACGFMGCTPGNPLACANAASLAIISGDMO QOVADIEVOLREQLAPARDAEMVADVLAIGVETTHPVNMAALQKFEVEGVNIR PFGKILVIMPPYITLPOLQRLTAANRAVODETFEQ"
2012..3052
/gene="biob"
2012..3052
/gene="biob"
/codon_start=1
/transl_table=1
/product="biotin synthetase"
/protein_id="AAA23515.1"
/db_xref="GI:145425"
/translation="MAHRPMTLSQVTELEFEKPLDILFEAAQVYHROHPROYOST LLSIKTGACPEDCKYCPQTSRYKGTGLAEERLMEVOYLESARAKAAGSTRFCGAM KNPHERDMPYLEQVGVGKAMGLEACMTLGTLSBSQQRANGLDYNNINLPTSPES YGNIIITRTYQERLDLTLEKVDAGIKVCSGIVGLGETVRDAGLILQILANLPTPES VPINMIVKVGCTPLADNDVDAPDFITIAVARIMPTSVYRSAGREGNNPQOALMC EMAGANSITFGCKLITTPNPEEDKDLDFKGLINPQZNAVLAGDNQOORLQALMT PTDLEYVNAAL"
3049..4203
/gene="bioc"
3049..4203
/gene="bioc"
/gene="bioc"
/codon_start=1
/transl_table=1
/product="7-keto-8-amino pelargonic acid synthetase"
/protein_id="AAA23516.1"
/db_xref="GI:145426"
/translation="MSMOEKINAAIDARRAADALRRRYPVAGAGRWLVADROYLNF SSNDYLGISHHPQIIRAMOGAEPFGSGSGSSGVSCSYVHQALEELAELEWIGYSAR LLFISGFANAGAVTAAMAKEDRIAOLBSLASILLESASLSPSLRRPANDYTHLGR LIASPCPOQAMVTEGVFSGMDGSAPIAEIQOVYQDHNGLWLVADDAHGTGVIGDHLG SCMLQKVELLVVTFGFGVSGAAVCSSTVADYLLQARHLIYSTSNPPAQAQL RASLAVIRSDGDAERKIALILTRFRAGVODLPFTLADSCSAIQPLIVDNSRAQL AKRLQOQSCWTAIRPPTVPAGTARLRLTLTAHEMDIDRLLEVLHGNC"
4190..4945
/gene="bioc"
4190..4945
/gene="bioc"
/gene="bioc"
/gene="bioc"

```

/codon_start=1
/transl_table=11
/protein_id="AA023517.1"
/db_xref="GI:145427"
/translation="MATVNNKQATIAAGRAAAHYEQHADLOROSADLLAMLPRKYT
HYLDAGCGPGMWSRHHRRRAQVATLDSPPMLVOAKODADHYLAGDIESLPLATA
TFDLAWSMLAVQWCMGKSTALRELRYRVRKGVAFITLVQGSILPERHQAQVDEHP
HANRFLPDEIEOSLNGVYQHHTIOPITLWFDLMSMRSLKIGATHLEGRDRL
TRSQLORLOLAMPQOGRPLTYHLFLGVARE"
4938..5597
/gene="biod"
/locus_tag="biod"
/locus="biod"
/codon_start=1
/transl_table=11
/product="dethiobiotin synthetase"
/protein_id="AA023518.1"
/db_xref="GI:145428"
/translation="MSKRYFVTGDTFVGKTVASCALLQAAKRAQYTAGYKRVASGS
KPEEGIRNSDALALQNRNSLQDLYATVNTTAEPTSPHISAGORPLESLWSAG
LRAEQADWYLVGAGGFTPLSDTETADWYQEQLPVILVGVKLCINHAMLTA
QVIOHAGLTLAGWVANDVTPPGKRHAQYMTTLRMIPRCWERSPGLOKIOKKRQES
T"
BASE COUNT 1363 a 1554 c 1631 g 1245 t
ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaatcgatccgtcaggtcag 25
Db 2110 GCATTTCATCTCTCGTCAGTGCAG 2134
|||||
RESULT 6
LOCUS A38246 5872 bp DNA
DEFINITION Sequence 1 from Patent WO9408023.
ACCESSION A38246
VERSION A38246.1 GI:2294844
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch O., Brass J., Fuhrmann M. and Shaw N.
TITLES BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 1 14-APR-1994;
LONZA AG (CH)
COMMENT
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
FEATURES
source
location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence="experimental"
23..28
/function="PROMOTER PTAC"
45..49
/standard_name="PROMOTER PTAC"
/evidence="experimental"

RBS 105..119
/standard_name="BIOC RBS NO.9"
/evidence="experimental"
117..1157
/gene="BIOC"
117..1157
/gene="BIOC"
/codon_start=1
/transl_table=11
/number=1
/evidence="experimental"
/product="BIOTIN SYNTHASE"
/protein_id="CA02325.1"
/db_xref="GI:229485"
/translation="MAHRPRWTLISQVTELEFKPLDLLFEAQVYHROHFDPROVOVST
LSTKIGACPECKYCPQSSRYKGTLEAEERLMEVROYLESARAKACSTPCGAM
KPHRDMPEYLEQWQVYKANGLEACMLTGTSQOARLNAGIDYNNHLDSPER
YCNITITTRTORLDTLEKVRACIKVCSGGIVGIGETVKORAGILLQLANLPDES
VFINMLVKKGIPLADNDVDVDFEIRITVAKRIIMPPTSVYLSAGREONQOAMC
FMAGANSIFYGCKLITPNPEBDKDLQLEFRKILGNPQOTAVIAGDNEQOORLEOALMT
PDTDEYVNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CA02326.1"
/db_xref="GI:229486"
/translation="MATVNNKQATIAAGRAAAHYEQHADLOROSADLLAMLPRKYT
HYLDAGCGPGMWSRHHRRRAQVATLDSPPMLVOAKODADHYLAGDIESLPLATA
TFDLAWSMLAVQWCMGKSTALRELRYRVRKGVAFITLVQGSILPERHQAQVDEHP
HANRFLPDEIEOSLNGVYQHHTIOPITLWFDLMSMRSLKIGATHLEGRDRL
TRSQLORLOLAMPQOGRPLTYHLFLGVARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF."
/ec_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence="experimental"
/product="DAPA SYNTHASE"
/protein_id="CA02327.1"
/db_xref="GI:229487"
/translation="MTTDDLAFDQRIHIMPTSTSPLYPVYVSAEGCELLISDGR
LVDMSSWMAIHGYNHNPOLNAAKSODIASHYVFGITTHAPAEICRLVATPDP
LECYFIADSGVAEVMKMAIOYQAKGEAROFITFRNGYHGDPEGASVDDPNS
MHSILKGYLPENLFAPARQSRMDENBEROVYGARLMAHREHITAIVTEPIYQAG
GMRATHPBWLKRIKICDRBSGLITLADILATGRTGKLRACERAEIAPDILGKAL
TGGTMTLSATLTTRVAVETISNGEAGCFMGPTFMGKPLCAANASLAILESGDWO
QVADIEVOLREOLAPARDAEWADRVIGAGIVVETHPVNNMALQKFEVQGWIRP
EGKLIYMPYIILPOOLRLTAIVNBAVODETFPCQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6

```

```

/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDLRGDKLPHHVTNMGCTGCDNISPLAMDVAGTK
SEVVCYDPDAPVGGMMHVVVNLPAOTRLPGCFSSGLVAMPDGLQTRDGTGK
YDGAAPPKGETHRYLFTVHALDIERIDVDEGASGAMGFFVHFSLASATFAMS"
5583..5605
stem_loop
5583..5644
terminator
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"

BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 gcattcgatcctcgatcgatcgac 25
        |||||||
Db      215 GCATTTCGATCCTCGTCAGGTGCAG 239

RESULT 7
A38251      5872 bp      DNA
LOCUS      Sequence 6 from Patent WO9408023.
DEFINITION A38251
ACCESSION A38251
VERSION A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS    Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL    Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT    Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694T 960227.
            Location/Qualifiers
FEATURES
source      1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PBO30A15-9"
            1141..1156
            /standard_name="BIOF RBS"
            1154..2308
            /gene="BIOF"
            1154..2308
            /gene="BIOF"
            /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
            /EC_number="2.3.1.47"
            /codon_start=1
            /transl_table=11
            /number=2
            /evidence=experimental
            /product="KAPA SYNTHASE"
            /protein_id="CAA02329.1"
            /db_xref="GI:2294850"
            /translation="MSMOEKINAAIDARAADALPRRYPVAGAGRMIVADROYLNF
            SSNDYLGSHHPQITRAWOOGAEQFGISGSGHYSGYSVHQALEELAEHLGSR
            LFTISGFANQAVIAAMAKEDRIADRLSHSLLEAASLSPSLRRFAHNDVTHLAR
            LLASPCGOQNVVTEGVFSMDGSDAPLAETIQVTOQHNGWIMVWDAGHGTGYIGEOGRG

```

```

SCMLQKVPPELLVVFEGKGGVSGAAVLCSTVADYLLOFARHLITYSTSPPAQOAL
RASIAVIRSDGEGARREKUALITRRAGVODLPTLIADSCSAIOPLIVGDNSRALQL
AEKIROOGCWVTAIRPPTVAGTARILRLTFAHEMQDIDRLLEVJLHGNG"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYEVTGTDEVEKTVASCALLQAAKAGRYTAGYRPVAGS
EKTPEGIRNSDALAQNSSLQIDYATVNPYTPAEPTSPHIIISQGRPIESLVMGAG
LRALQOQADWLVLEGAGGWEPTPLSDFTTFADVATQOLPVLVGVKGLGINHAMLTA
OVIQHNAGLTLAGVANDVTPPGKRHAHEYMTTLIRMIAPLLGRIPLALANPENAAATGK
YINLAEVDASTLFTSRL"

BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 gcattcgatcctcgatcgatcgac 25
        |||||||
Db      215 GCATTTCGATCCTCGTCAGGTGCAG 239

RESULT 8
A93674      5872 bp      DNA
LOCUS      Sequence 1 from Patent EP0798384.
DEFINITION A93674
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS    Birch,O. and Brass,J.
TITLE      Biotechnological method of producing biotin
JOURNAL    Patent: EP 0798384-A 1 01-Oct-1997;
            LONZA AG (CH)
COMMENT    Location/Qualifiers
FEATURES
source      1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PBO30A-15/9"
            1..96
            /function="PROMOTER PTAC"
            /evidence=experimental
            23..28
            /standard_name="PROMOTER PTAC"
            105..119
            /evidence=experimental
            117..1157
            /gene="BIOB"
            117..1157
            /gene="BIOB"
            /codon_start=1

```

```

/transl_table=11
/evidence-experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPWTLSQVTELEFRLDILFEAOVHROHFDPROVOST
LILIKTGACPEPCCKYCPOSSRYKTGLEAEERIMEVEQVLESARKAKAGSTRFCGAAW
KNPHERDMPIYLEQWVGKAKGLEACMLTGLTSGSQAKRLANAGLDYNNHNLDTSPER
YGNITTRTYERLDTELEKVDAGIKVCSGIVIGETGVDRAGLILQLANLPTPES
VPIMLVVKCTPLADNDVDADFIRITIAVARIMPTSVRLSAGREONNEOTQAC
FMAGANSIFYCKLITTPNPEKDLQIFRLGLNPQOTAVLAGDNQOQRLQALMT
PDTEYVNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVYKQATIAAFGRRAAHYEHADLOROSADALLMLPORKYT
HVLACGPGMGRHMRERHQAQVATLDSPEMLVQAKADHYLAGDIESLPLATA
TFDLAMSLAVQKCGNSTALRELIRVRRGVAFITLVQGSLEPILQWQVADERP
HANKFLPPEDELSNGVHYOHITQPTLTFMEDDLSMRSLKIGIGATHLEGDRPRL
TRSQLORQLAMPQOGRPYTLHLFLGVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANANATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAPDORHIMHPTSWTSPYVYVSAAGCELLISDRR
LYDSSSMMAIHGYNNHLPOLNAAMKSOIDAMSHVMRGITHAIEICRVLAMTPOP
LECYFLADSSVAYEVAMKALQYQAKGARKGFLTFRRNGYHGDFTFGAASVCDPNS
MHSIMKGYLPENLEAPAPQSRMDGEMDERMVGARLMAHREHIAVITPEIVQAG
GMRVYHEMLKRIKICDRGILLADELATGFRGKLFACAEIAPIELICIGAL
TGGTMTLSATLTREVAETISNGEAGCFMHGPTFMGNPLACAANASLALFEGCDMO
OVADIEVQLREQLAPARDAEMVADVRLGAGVETTHPVNMAALQKFVEQGVMLRP
FGKLIIMPYIILPQOLQRLTAVNNAVODETFPCO"
5088..5100
/standard_name="ORF1 RBS"
5098..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGLKPLRHVFNMGVYDGNISPLHAMDVPACTK
SFVVTCTDPDAPGSGMMWVNVNLPADTRVLPQCGSGIAMPDGVLOTTRDFGKTG
YDGAAPKGETTHYITVALDIERIDVDEGASGAVGVNHHSLASASTAMPS"
5583..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1318 a, 1352 c 1695 g 1307 t

```

```

ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.071;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatctgcattccgtcaagtcag 25
Db 215 GCATTGCATCCGTCGATGAGTGCAG 239
RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
AUTHORS Birch, O. and Brass, J.
TITLES Biotechnological method of producing biotin
JOURNAL Patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
Location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANANATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSWQEKINAAIDARRADALRRRYPVAGAGRMVLVADPROYLN
FLEFISGRANQAVIAAMAKEDRIADRLSHASLSLEASLSPPSOLRRFAHNDVTHLRA
LIASPCPGQOMVTEGVFSMDGSAPLAEIQQVTOQHNGMLMVDADHGTGIVGQKGS
SCMLOKVRBELLVTFEGKFGVSGAAYCSSTVADYLIQFARHLIYTSMPAPAOAL
RSILAVIRSDSDGARREKIALITFRFAGVODLPFTLADSCSAIOPITVIGNSRALQL
AEKLRQCGCWTAIRPVPVAGTARLRLITLAHEMDIDRLLEVLHNG"
3030..3045
/standard_name="BIOD RBS"
3043..3753
/gene="BIOD"
3043..3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="D7B SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYEVTGDTDEVGKTVASCALLQAKAAGYTPAGYKPYAGSS
EKTEPGLNSDALAQNNSSLDQDYATVAVNPYTPAEPPSPHILISAQEBRPIESLVMSG
LRALEQADWNLVBGAGCMFTPLSDTFFADWYVQEOPLVILVGVKLCGINHMLTA
QVIOHAGLTLAGWAVANDVTPGKRHAETMTLTBMIPADPLIGELIPWLAENPENATCK

```

BASE COUNT

1318 a, 1352 c 1695 g 1307 t

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 YINLAFFVASTLGFSTRSL"
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.071;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatttcgacccctgcaggtgcag 25
 |||||||
 Db 215 GCATTTCGATCCTCGTCAGGTGCAG 239

RESULT 10
 ARI01809 5872 bp DNA PAT 14-FEB-2001
 LOCUS
 DEFINITION Sequence 1 from patent US 6083712.
 ACCESSION ARI01809
 VERSION ARI01809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000.
 FEATURES
 Location/Qualifiers
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.071;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatttcgacccctgcaggtgcag 25
 |||||||
 Db 215 GCATTTCGATCCTCGTCAGGTGCAG 239

RESULT 11
 ARI01810 5872 bp DNA PAT 14-FEB-2001
 LOCUS
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION ARI01810
 VERSION ARI01810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000.
 FEATURES
 Location/Qualifiers
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.071;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatttcgacccctgcaggtgcag 25
 |||||||
 Db 215 GCATTTCGATCCTCGTCAGGTGCAG 239

RESULT 12
 AE000180 11022 bp DNA BCT 01-DEC-2000
 LOCUS
 DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete
 genome.
 ACCESSION AE000180 U000096
 VERSION AE000180.1 GI:1786988

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Escherichia coli K12.
 Escherichia coli K12.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J.J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mau, B. and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503

REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett, G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

COMMENT
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
 30332 (e-mail: mark@amr.gatech.edu). Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 site nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (http://cgsc.biology.yale.edu). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"

```

repeat_region /db_xref="taxon:83333"
               <1..130
               /note="REP (repetitive extragenic palindromic) element;
               contains 4 REP sequences"
gene           complement(147..1430)
               /gene="ybhc"
               /note="b0772"
               complement(147..1430)
               /gene="ybhc"
               /function="orf: Not classified"
               /note="f427: 98 pct identical to fragment YBHC_ECOLI
               SW:P46130 (300 aa) but contains 127 additional C-terminal
               residues"
               /codon_start=1
               /transl_table=1
               /product="putative pectinesterase"
               /protein_id="AAC73859.1"
               /db_xref="GI:1786989"
               /translation="MNTFSVRLALAFGVTLTACSSPPDQPSDQAPGSSRPI
               LSAEQAQNPQAHYFASLTGGAAMNPSPTLPAQDPFVYVAGTGTHTTIOAIVD
               AAIKRTNKROYIAVMPGEYQIVYPAAPGGITLGTGKPIDVAKIGLSLQDGMSPA
               DWRHVDVPRKGYMPKPAWYDSCQSRSDSGVCSAVFWSQNNGLQNLNLTENT
               LGSDVDAGNHPAVALRTDGDQVOIINNVLIGRQNTFVNSGVQNLRTNRPRTIVT
               NSYIEGDVIDVSGGAVFNDTEFRVYNSRTOOEAVYFAPATLSNIYYGLAVNSRP
               AFGDVVALGRSLDVAQNTGVYIRDSAINEGENIAKPAWADAVISNRPAGNTGSVD
               DNEIORNLNDTNNRMVEYNNRGVSKVVAEAKK"
promoter       complement(1507..1535)
               /note="Factor Sigma70; predicted +1 start at 806574"
               complement(1582..2058)
               /gene="ybhb"
               /note="b0773"
               complement(1582..2058)
               /gene="ybhb"
               /function="orf: Unknown"
               /note="f158: 99 pct identical to YBHB_ECOLI SW: P12994"
               /codon_start=1
               /transl_table=1
               /product="orf, hypothetical protein"
               /protein_id="AAC73860.1"
               /db_xref="GI:1786990"
               /translation="MKLIISDLRDKRLPHRHVFGNGMGYDGDNI SPHLAMDVPAGTK
               SFVYCTDPDAPITGSGMHWVYVNLPAIDRVLPQGGSGI.VAMPDGVLTQTRDFGRTG
               YDGAAPKGETHRYIFTVALDIERIDVEGASGAVGCVNHFHSLASITAMFS"
               complement(2108..2124)
               /note="central position to predicted promoter:85..5"
               /bound_moiety="Rhas predicted site"
               complement(2117..3406)
               /gene="b10A"
               /note="b0774"
               complement(2117..3406)
               /gene="b10A"
               /EC_number="2.6.1.62"
               /function="enzyme; Biosynthesis of cofactors, carriers:
               Biotin"
               /note="f429: 100 pct identical to B10A_ECOLI SW: P12995"
               /codon_start=1
               /transl_table=1
               /product="7, 8-diaminopelargonic acid synthetase"
               /protein_id="AAC73861.1"
               /db_xref="GI:1786991"
               /translation="MTTDLAFDRHITHWPIYTSWSPLEFVYVVSAAEGCEILLSDGR
               LVDGMSMAAIIHGVNHPOLINAAKSQIDAMSHVMFGITNAPELICKRIYAMPPOP
               LECVPLADSGSAVEVAMKALQYQAKQEARQRLPTFNGYHGDTEFGAMSYCDSDNS
               MSHIMKGYLPENLEFAPQSRMDGEMDERDNGFALYMAHRETAIAVLIETVGGAG
               GMRMYHPWIKRIKICDREGILLADIELATGEGRGKLFACENHAIVDILICGLAK
               TGGTMTLSATLTTRVAELETISNGEAGCGEFTFGNGLPACAAANASLAIIESGWMQO
               OYADIEVOLREOLAPARDAEMVADRYVCAIGVETTRVNNALQKFFVDOGVNIRP
               FCKLITLMPPIITLIPQOLRIYAAVNAVQDEFECQ"
               complement(2193..2221)
               /gene="b10A"
               /note="factor sigma70; predicted +1 start at 807260"
               3411..3450
               /note="central position to bioB promoter: -20"

```

```

protein_bind  /bound_moiety="BioB documented site"
               3411..3450
               /note="central position to predicted promoter: -20; genetic
               evidence for the site"
protein_bind  /bound_moiety="BioB predicted site"
               complement(3411..3450)
               /note="central position to predicted promoter:50"
               /bound_moiety="BioB predicted site"
               complement(3411..3450)
               /note="central position to bioA promoter:50"
               3413..3441
               /bound_moiety="BioB documented site"
               /note="factor sigma70; promoter bioB; documented +1 at
               808525"
               complement(3447..3473)
               /note="factor sigma70; promoter bioA; documented +1 at
               808515"
               3493..4533
               /gene="b10B"
               /note="b0775"
               3493..4533
               /gene="b10B"
               /EC_number="2.8.1.-"
               /function="enzyme; Biosynthesis of cofactors, carriers:
               Biotin"
               /note="c346: 99 pct identical to B10B_ECOLI SW: P12996"
               /codon_start=1
               /transl_table=1
               /product="biotin synthesis, sulfur insertion?"
               /protein_id="AAC73862.1"
               /db_xref="GI:1786992"
               /translation="MAHRPMVTLSCVTELEFKPLDILFEAQVYRQHFDPQVQVST
               LLSIKVGAPEDECKYCPQSSRYKTGLEMERLMEVEQVYESARKKAAGSTFEFGAAM
               KNPHRMDPYLEOMVQVQKAMGLEACMTLGLTSSQARLANAGLIDYNNHLDTPPEF
               YGNITRTYQERLDTLEKVDAGIKVQSGSIGVGETVGRAGLIDYNNHLDTPPEF
               VPINLVKVGKPLADNDVDAPDFIRTIYAVARIIMPVSYRISAGREOMNEOTQAMC
               FMAGNSIFCYGCKLLTTPNPREDKDLQFRKGLINPQOTAVIADNEDQORLEQALMT
               PPTDYVNAAL"
               4530..5684
               /gene="b10F"
               /note="b0776"
               4530..5684
               /gene="b10F"
               /EC_number="2.3.1.47"
               /function="enzyme; Biosynthesis of cofactors, carriers:
               Biotin"
               /note="c384: 100 pct identical to B10F_ECOLI SW: P12998"
               /codon_start=1
               /transl_table=1
               /product="8-amino-7-oxononanoate synthase"
               /protein_id="AAC73863.1"
               /db_xref="GI:1786993"
               /translation="MSWQETINALLDARRADALRRRYPYAOGAGRWLVADDRQYINF
               SSNDYLSLGHSHPOITRAMQGAQEQFGSGSGHVSQSVYVHQAEELEAEWGYSLA
               LLFISGFANQAVIAAMAKEDRIYAADRLSHASLLEASLSPQLRFAINDVYHLAR
               LLAPRCQGOQVVTGEGVSMGDSAPLAETIQVYQOHNGMLAMDADAGTGTGEOGCG
               SCMLQKVKPELVYTFKRGVSGAAVLCSYVADYLLQPARHLIYSTSNPRAQOAHL

```

Query Match 100.0% Score 25; DB 1; Length 11022;
 Best Local Similarity 100.0%; Pred. No. 0.073;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgcacccgcgcagcgcag 25
 |||||||
 Db 3591 GCATTTCATCTCGTCAGGTCAG 3615

RESULT 13
 AE005258 13501 bp DNA BCF 21-MAR-2001
 LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
 DEFINITION of 155
 ACCESSION AE005258 AE005174

```

VERSION      AE005258.1  GI:12513751
KEYWORDS
SOURCE       Escherichia coli O157:H7 EDL933.
ORGANISM     Escherichia coli O157:H7 EDL933
              Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE    1 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
              Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
              Nature 409 (6819), 529-533 (2001)
JOURNAL      MEDLINE
PUBMED       11206551
REFERENCE    2 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
              Direct Submission
              Submitted (22-OCT-2000) Laboratory of Genetics, University of
              Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
              Location/Qualifiers
              1. 13501
                 /organism="Escherichia coli O157:H7 EDL933"
                 /strain="EDL933"
                 /serotype="O157:H7"
                 /db_xref="taxon:155864"
                 /note="Enterohaemorrhagic"
                 <1..7576
                 /note="0-island #36: Region of the EDL933 chromosome not
                 homologous to E. coli K-12 MG1655: Cryptic prophage
                 CP-933K: includes one copy of the 13 bp direct repeat that
                 flanks the prophage"
                 66..665
                 /gene="lomK"
                 /note="lomK"
                 /db_xref="GI:12513752"
                 /function="putative membrane: Other or unknown (Phage or
                 Prophage Related)"
                 /note="Residues 1 to 199 of 199 are 72.86 pct identical to
                 residues 1 to 199 of 199 from Genpept 118 :
                 g117532789|gb|AAFG3231.1|AF151091_2 (AF151091) lom
                 [prophage P-E1bA]"
                 /codon_start=1
                 /transl_table=11
                 /product="putative outer membrane protein of prophage
                 CP-933K"
                 /db_xref="GI:12513752"
                 /translation="MKRCALILSLATCLAVSPVPMASEHOSILSACYLHASTDARE
                 SDDLNGIVKTYRFTDTLGLTSTFVNADEOKTHYSDTRMHDVYRNKFSVMAG
                 PSRVNWFSAVAMAGVAYSRVSTFSGDYFRTVDMKRTHDVLGSDADARYSNPSLAW
                 GAGYQFNPTESAVADVAYEXSGSDMRTDFIVGVYKF"
                 719..2041
                 /gene="20982"
                 /note="20982"
                 /function="putative structure; Structural component (Phage
                 or Prophage Related)"
                 /note="Residues 164 to 440 of 440 are 68.79 pct identical
                 to residues 381 to 645 of 645 from Genpept 118 :
                 g145854361|gb|AAD25464.1|AF125520_59 (AF125520) putative
                 tail fiber protein [bacteriophage 933W]"
                 /codon_start=1
                 /transl_table=11
                 /product="putative tail component of prophage CP-933K"

gene
CDS
        /protein_id="AAG55138.1"
        /db_xref="GI:12513753"
        /translation="MGMAVQISGVLYKDGACQPIQNTQIKARRNRYVYVNTYASE
        NPDEAGRYSMDEYEGYSVTLIVEGPPSPHAGTISYEDSOPTINDLFGATVEDAR
        PEALRFREDEWAEBAARHAEKAKNAGEASTSARNAGISAKAEANADNYSABEASE
        SARQAAESASAKSEASESSSSAKAEASISLOSATDAELSKTAESANADARDA
        TSTTEKARSASAEASQOSRIAEADVIRPIYVPGPPKGGPGAGPGQGGKAGPA
        KDTGPGAGATNGERGPGDTPGAGPGQPKDGRERGEGTGTGNAGPGPGPGDCAKPA
        PGQPGGEGTAACPGVATGPGQGGPGQGGTQTRFRIGPRITITNSYCFPGTDCALI
        TGTFTLDPDATTQVGMFQHLQVRFQDGMQDVKGIDEVSDTGRTE"
        2004..2312
        /gene="20984"
        /note="20984"
        /function="orf; Other or unknown (Phage or Prophage
        Related)"
        /note="Residues 1 to 102 of 102 are 98.03 pct identical to
        residues 28 to 129 of 129 from Genpept 118 :
        g145854371|gb|AAD25465.1|AF125520_60 (AF125520)
        hypothetical protein [bacteriophage 933W]"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933K"
        /protein_id="AAG55139.1"
        /db_xref="GI:12513754"
        /translation="MKRAVTOAELENDNMNIRKLMQSLGCGKHDCNGRSUTQALR
        LGPADLLESDENQITIPQDRVITTYVYILDADKKQIQCYVRPLQILRADGTWENIGMK"
        2489..3469
        /gene="20985"
        /note="20985"
        /function="orf; Other or unknown (Phage or Prophage
        Related)"
        /note="Residues 13 to 326 of 326 are 55.73 pct identical
        to residues 18 to 331 of 336 from Genpept 118 :
        g116603671|gb|AAFG3327.1| (AF170176) hypothetical protein
        predicted by Glimmer [Salmonella typhimurium LT2]"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933K"
        /protein_id="AAG55140.1"
        /db_xref="GI:12513755"
        /translation="MLSPIKTPPHNSVNIQVSPCOTVSPAGKEYLEKVIDETPLIF
        QMEPNPERKKNQDEVPVYNKKQPRYLDNTNARISDSKIGIFVNDGDSVYNKAFS
        KLEKDEBNWITITREDPDSMTDRKLSLDIYHDTICORLTERKDEYLLNLEKELR
        EISKADQSLISMVAKRKNHAFPFENLALKAGELFCKYTNKRNHGISFGECITLD
        MDMLITGKLTGIYAPDGISMNVDRNDVSIENSATIVNSNHPALLEGLSPMHSKVD
        AHPYDGLGKGVKKYFNFTPLHYNHCFIEFNHNHIIIMT5QYTCSSM"
        3530..4522
        /gene="20986"
        /note="20986"
        /function="orf; Other or unknown (Phage or Prophage
        Related)"
        /note="No significant matches"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933K"
        /protein_id="AAG55141.1"
        /db_xref="GI:12513756"
        /translation="MKIPSLQSNPNFNSAPAGYCAPLAPNRAENAVADYVLDIGKRIPL
        SAADLSNVESYIRAVYDSDSRLLDHTYVMIGNTYVLDLSRSCQTRDADVSGITNEK
        VHTGCLIRNVEYELNESSVKIDTIDQSLTQENLETFDVGDEPLTFPICGEBNDNEPY
        VSPFVAPDQSYEMEPQWQGLIHETLIHHVGGSSDPGSDNSETLGPTEILARVADQLG
        WSPVDFGVAPDEPEALRLRLNALROAAMRHEENRAFEERLGTISRYEASPDPT
        EYSAVSNIGYGFQIDHDFGLAINDNLQDANOQLQVHGAPYITFGDVVKNHOQ"
        5350..6231
        /gene="20989"
        /note="20989"
        /function="orf; Other or unknown (Phage or Prophage
        Related)"

```

```

/Note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AA655142.1"
/db_xref="GI:12513757"
/translation="MLSPSYNLCGSMNSLTRNLTPSDNRYLSSVRDAVAHSDNGACV
KVGNTKTRVATDNKFCVTRSHSGCTNLHLRGLGWKGETSRKTEVLMASVPASAM
ERGLVSHNRDPLPVDAAPLPSVDNRLSPGNVIGKGNNAVYEDADADATKVLKM
FTTSQSEVETSEVRCNOYAGASAEKIGNNDIGIMDKINGESLLNISLPAC
AEHAIDMFDRLEOKGILFVDTETNVLVBRANKNEFPNIDISSYNSVSDRSMSQIMQ
SYHGKDLISVLSVKI"
6462..7160
/gene="20990"
/length="7160"
/gene="20990"
/function="orf: Other or unknown (Phage or Prophage
Related)"
/Note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AA655143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNIVLHOSRSMSPDIESIYKASVQWTKNNPOLRF
OGPDNIYXOEALADKIGSETGRVLLNLESISRLKSETVYIHLNSSLGVMARHD
IDANHRGTGSDRFNCINAVBYPCGEGISVADPHATTVFHELHVFPHNLGEMLVES
SRASOKYSPLLEEARVIGIGAFSEEVLSENKFHEIGMPRTSYPSXSLIHDNNI
VSLGQVRLHPLL"
complement(7668..8144)
/gene="ybhb"
/Note="20992"
complement(7668..8144)
/gene="ybhb"
/function="orf: Unknown function"
/Note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/transl_table=1
/product="orf: hypothetical protein"
/protein_id="AA655144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGLKPLRHRVFNKGYGDGNIISPLIADVDVPAKX
SFVYTCYDPAPATSGMMHWVYVLPADTRVLVPOGEGSLVAMPDGLIOTRDPKGTG
YDGAAPKKGEMHRIYVHALDYERIDVDEGASGAMVGFNHSHLASATAMPS"
complement(8203..9492)
/gene="b10a"
/Note="20993"
complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.074;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaattgcattccgtcagatgcag 25
|||||
|||||
|||||
|||||
|||||
Db 9677 GCATTGCATTCCTCGTCAGTGCAG 9701

RESULT 14
AP002553 297816 bp DNA BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
DEFINITION AP002553 BA000007
ACCESSION AP002553.1 GI:13360211
VERSION
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (sites),

```

```

AUTHORS Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, T., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C. and Shinagawa, H.
TITLE Complete nucleotide sequence of the prophage vY1-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
JOURNAL MEDLINE
REFERENCE
AUTHORS Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
TITLE Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20573356
JOURNAL MEDLINE
REFERENCE
AUTHORS 3 (sites)
Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shinagawa, H.
TITLE Complete genome sequence of the prophage vY1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
JOURNAL MEDLINE
REFERENCE
AUTHORS 4 (sites)
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsuno, E., Nakayama, K., Murata, T.,
Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shinagawa, H.
TITLE Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
JOURNAL MEDLINE
REFERENCE
AUTHORS 5 (bases 1 to 297816)
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
Hayashi, T.
TITLE Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-info.osaka-u.ac.jp; Tel: 81-6-6879-8365,
URL: http://www.gen-info.osaka-u.ac.jp/; Fax: 81-6-6879-2047)
genome project.
COMMENT
FEATURES
source
location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="Ecs0753"
79..1245
/gene="Ecs0753"
/Note="similar to SUCC_ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=1
/evidence="not experimental"
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/translation="MNLHETQAKQLPARYGLPAVVGACTTPREAEASAKIGAGPVV
VKQYHAGRGKAGGVKVVNSKEDI RAEEMWLKRLVTVOTDANQGVNDILVEAT
DIAKELVIGAVVDSRRVVRPMASREGVEIEKVAEETPHILHRAVALDPLTGPPYOG
RELAKELIEGKLVQOQFPIKMGATLIEERLDLALIEINPLVITRKODLIDCKSLGA
DGNALFROPDLREKMDQSDQEPREKQAAQWELVVALDGNITGCVNAGLAMGMIDIV
KLHGEPANPLIDVGGGAKREKRVTAFTIISDDKVKAVIVNIPGCIYRCDLIANCIIG
AAVEYGVAVPVVRLGNNALGAKKLADSGNLIIIAKGLTDAQGVAAVEGR"

```

gene 1245..2114
/gene="ECs0754"
CDS 1245..2114
/note="similar to SUCD_ECOLI g11786949 percent identity 100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSILIDKNTKVCQGFPGSGCTFHSQDAIYVGTMGVTPGK
GTHIGLPYNTVREAVATGATASVITVPAPPCDSILEAIDAGIKILITTEGIP
LDMLTVKVLDEAGVIRMGPCVITPGECKIGIPGHIRPGKGVISRSGLTTE
AVKQTDYGFQSGTGVIGDPTPGSNFIDILEMEKDPQTEALVIMIGEISGSEEA
AAVIEHVTKEPVVYIAGVTAPKGRMGHAGAIAGSGTADKEFALEAGVTVRS
LADIGELKTVLK"
gene 2781..3686
/gene="ECs0755"
CDS 2781..3686
/note="probable transcriptional regulator/similar to transcriptional regulators e.g. glycine cleavage system Escherichia coli g1417043|sp|P92004|CQVA_ECOLI percent identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELLVTEEVARHSEYTRAAEELATGSAVPROVAL
BEFLRPLFESHKRIPLNDGKYIGVETLNLERDNTIMTQPTVOYIELAVN
PFTSTHMLPNIHEFTKLHPDIIVNIHSLANGDPLNREDAVIMRENFCAVAFVY
LPEETILPVCSGLSLAMSDOKSLVALLTELPLHSTRTGMEEMFALSGVSLVN
NGPRDLISMLIAAVRSNLPRFALQHDLDGDMVITPCDVPITRGNRIIMTQE
EKSDSPHLOOFREMLAKSVPEM"
gene complement(3720..4322)
/gene="ECs0756"
CDS complement(3720..4322)
/note="probable cob(1)alamin adenosyltransferase, similar to cob(1)alamin adenosyltransferases (corrinoid adenosyltransferases) e.g. [Escherichia coli] g1115148|sp|P13040|BYUR_ECOLI percent identity 67 in 200 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative cob(1)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTEVRHOROOLKLEQVDTVAALIEKKGILVFTGNGK
KSTAGVYTRAVGCHCKYCAVYIKGMDNGEYNLQIGVEPHIMGSGFWENOR
QADIDAAKEVWSESKRLADKRDVLYDELTYMLAYHLDIEVYIASLQNPACQSV
IYTRGCHSQILKMAUTVSEIRPVKHAIFDNGIOADPGIDW"
gene complement(4332..5984)
/gene="ECs0757"
CDS complement(4332..5984)
/note="probable fumarate hydratase, similar to fumarate hydratases e.g. fumarate hydratase class I, aerobic [Escherichia coli] g1120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545 aa"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"

/translation="MSKPIFMOEPFLONKGTETYLISDQHTITELDEEVIKIAPE
ALTLIARAFVFAFPFLRSNHLCOVASIISINDOASQENNLALOLLRNVEYAKGLP
NCOQDTGATTVASKGQOITRGNDALSKSGITSPQENNLRSQNAADVTYVNTQ
TNLPADIDISAVAGDEHFLCVNKGGSARKALVQETSLAQPEKTRFLIEKKSLS
GTACAPRYHIAFVYGGISADQTLKIAKLASTKYDNLPTSGNDQGAFDILEYVL
EASQPCIGIAQFGCGKYFADIRVIRLPRHGGSCPIMALSCADRNIKIKHGLTWL
EKLEHNCGOYIPASLRENNHQAQVODLNDPLMODLARLPGVTRVLSGPIVAVR
DIANHKIKARLDGEPMEPEYLRKHIIYVYGPAKTPENMGCSLGPPTGGMQYIDTF
QAAQCSLYMLSKNRSQVTDACHKRGFGTSGGAAALAQEYKSLRCLEYBELG
MEAVMMHEVENDLAEFLVLDKGNMFSSQEPDQHRACAPRAGH"
gene complement(6092..7372)
/gene="ECs0758"
CDS complement(6092..7372)
/note="probable transport protein, similar to glutamate/aspartate transport proteins (proton glutamate symport proteins) e.g. [Bacillus stearothermophilus] g1121467|sp|P24943|GLUT_BACST percent identity 38 in 416 aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLTPLMLALVGMIIQVYVNTASPEAKLYAOEISFT
TTPRLIKMTIAPLVSTLYVGIKMGDAKALRIPSKTLFLFCASLISALGILTV
NEFMPGTGINFVHAGATTTGVAAEPTLTVFSHAPPTSIVYMANHETLOVPSI
FLGCSLIAIGKGSATVIALHSDLSAHAKLITGVMDLPRVIRALISALIAEGGLAV
VSAGITFMGEFEFTMLLMVLLIGLAIYVPCIRRLRALSEPALAFETTSSEAAV
GLTKLILQFGVSPKIASFVLPIGTGFNMGSMGSPATVAFIPAOANILHISIEQITM
LLIILKTSKMGAVPRASMYVIAATLQFNIPENGLILLGVPFLDIMGSRATNVMSN
AMGAAMVSRMGHEHGEGRGKALKPESVVALP"
gene complement(7533..7850)
/gene="ECs0759"
CDS complement(7533..7850)
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"
Query Match 100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.083;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gcatctcgaatcctgcagtgacg 25
DB 92409 GCATTTCGATCCTCGTCAGTGCAG 92433
RESULT 15
AF250776
LOCUS 5526 bp DNA BCT 31-JAN-2001
DEFINITION Uncultured bacterium pCosHE2, hypothetical 17.1 kDa protein in
modc-bioA intergenic region, DAPA-aminotransferase BioA (bioA),
biotin synthase BioB (bioB), KAPA synthetase BioF (bioF), and
biotin biosynthesis protein BioC (bioC) genes, complete cds; and
dehbiobiotin synthetase BioD (bioD) gene, partial cds.
ACCESSION AF250776
VERSION AF250776.1 GI:12620124
KEYWORDS uncultured bacterium pCosHE2.
SOURCE uncultured bacterium pCosHE2.
ORGANISM Bacteria; environmental samples.
REFERENCE 1 (bases 1 to 5526)
AUTHORS Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streif, W.R.
TITLE Direct cloning from enrichment cultures, a reliable strategy for
isolation of complete operons and genes from microbial consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE 20575196

PUBMED 11133432
 REFERENCE 2 (bases 1 to 5526)
 AUTHORS Entcheva, P., Liebl, W. and Streitz, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
 FEATURES
 source
 1..5526
 /organism="uncultured bacterium pCosHE2"
 /db_xref="taxon:143797"
 /clone="pCosHE2"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(52..528)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical 17.1 kDa protein in modC-bioA
 intergenic region"
 /protein_id="AAG60577.1"
 /db_xref="GI:12620125"
 /translation="MKLISNDLRDGLPHRHVPMNGYDGDNTSPHLAMDVPAGTK
 SFVVCYDPAATGSGMMHWYVLPADTRVLPQSGSLVAMPDGLVQTRDCEKAG
 YDGAAPKGEHRYFTVHALDIERIDVDEGASAMGFVNHSLASASITAMFS"
 complement(587..1876)
 /gene="bioA"
 /complement(587..1876)
 /gene="bioA"
 /note="7,8-diaminopelargonic acid
 synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase BioA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDDLAFDQRIHMPYTSSTPLPYVVSAGCELLTSDGR
 LVDMSSMMAIHGYNHLPOLNAAMKSIDAMSVMEFGITHAHAEICRKLVTMPQ
 LKCVFLADSGSAVEAMKALQYMOAGKAPRLTFRNGAGDIEGSLAMSCDPS
 MSLMKGYLPENLEAPAPQSMGEMDERPMVSPALMAHRIHIAVITEPIVQAG
 GGRMYHPEMLKRIKICDRGILLIADETATGTRGKLPACHEAETAPDILGRL
 TGGTMTSATLITREVAETISNEAGCFMGPTFMGNPLCAANASLALLESQWOC
 QVADIEVQLREOLAPADAEVADVAVLGAIVETHPVMAALOKFVEQGVWIRP
 FGLIYLMPEYIILPOLRLTAANRAVODETFEQC"
 1963..3003
 /gene="bioB"
 1963..3003
 /gene="bioB"
 /gene="bioB"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase BioB"
 /protein_id="AAG60579.1"
 /db_xref="GI:12620127"
 /translation="MAHRPRWLSQVTELPFERLIDLLFEAOVVRHNDPRVOYST
 LLSITGACPEDECKKCCQSSRYTKGLAEELMVEYVLESARKAKAGSTRPCWAM
 KNREHMPYLEBOWGVKDLGLACHTLSTESQARLANGLDYNNHNDTSEF
 YGNITTRYQERLDLEKVDAGIGVCSGIVGLGETYRAGLLIOLANLPTPES
 VPINMLVKVGTPLANDVDADPFRITLAVARIMPTSVRLSAGREOMEDTQAC
 FMAGANSIFYGCKLITTPNEEDKDLTFKTLGILNPQTAVIAGDNEQOORLEQALMT
 PTDVEYNAAL"
 3000..4154
 /gene="bioF"
 3000..4154
 /gene="bioF"
 /gene="bioF"
 /note="8-oxononanoate synthase"
 /codon_start=1
 /transl_table=11
 /product="KAPA synthetase BioF"
 /protein_id="AAG60580.1"
 /db_xref="GI:12620128"
 /translation="MSWQKINAALDARADALRRRYDVAOGAGKMLVADROYLNF
 SSMDYGLSHHPQIIIRAMOGAEOFGIGSGSHVGSVYHQALEELAEWLGYSA
 LFIISGFANQAVIAAMAKEDRIADRLSHLSLEASLSPSOLRRFHHNVTIHLAR
 LIASPCGOOLVYTEGVSMDDGSAPLAEIQVTOOHNGMLVDDAAGHGVIGEGGRG

gene
 4141..4896
 /gene="bioC"
 4141..4896
 /gene="bioC"
 /note="reaction step prior to synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=11
 /product="biotin biosynthesis protein BioC"
 /protein_id="AAG60581.1"
 /db_xref="GI:12620129"
 /translation="NATVKAQIAAFGRALRYEHOADIOROSADVLAMPERKYT
 HYLDAGCGGMMTRHMRERAOVTALDLSPPYLVAROKDADHYLADISPLATA
 TTDLANSLAYVOCNGLSTALRELYRVRSKGVAFYLLVQSLPELQAWQADVERP
 HANRLPDEIEOSLVNGVYOHIOPTITLMPDALSAMKSLKIGATHLHGGRPRL
 TRSOLQRLAMPQOGRYPYVHFLVIARE"
 4883..>5526
 /gene="bioD"
 4883..>5526
 /gene="bioD"
 /note="DTP synthetase"
 /codon_start=1
 /transl_table=11
 /product="dethiobiotin synthetase BioD"
 /protein_id="AAG60582.1"
 /db_xref="GI:12620130"
 /translation="MLVSKRYVPTGTDTREKTVASCALLQAAKAGRYRAGYRVAS
 GSEKPEGRNSDALALRNSSLQDIATVNPYTPAETPSHIIISAGRTLESSVMS
 SGRLAEQADWLVGAGGWTPLSDFTFADWTOLOPLVIVGKLCINHAML
 TAOAIQHGGLTLGAVANDVTPPGKRAHEVITTLTRMPALDELIPWLAENPE"
 1507 c 1567 g 1178 t

BASE COUNT 1274 a 1507 c 1567 g 1178 t
 ORIGIN

Query Match 93.6%; Score 23.4; DB 2; Length 5526;
 Best Local Similarity 96.0%; Pred. No. 0.46;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gcatcgcattccctgcagtgag 25
 |||||||
 Db 2061 GCATTTCGATTCCTCGCAGTCCAG 2085

RESULT 16
 AF250768
 LOCUS 4000 bp DNA BCT 31-JAN-2001
 DEFINITION Uncultured bacterium pCosFS1 BioA-like protein (bioA) gene, partial
 cds: BioB-like protein (bioB) and KAPA-synthetase (bioF) genes,
 complete cds; and biotin biosynthesis protein BioC (bioC) gene,
 partial cds.
 AF250768
 AF250768.1 GI:12620095
 VERSION 1
 KEYWORDS
 SOURCE uncultured bacterium pCosFS1.
 ORGANISM uncultured bacterium pCosFS1.
 Bacteria; environmental samples.
 REFERENCE 1 (bases 1 to 4000)
 AUTHORS Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streitz, W.R.
 TITLE Direct cloning from enrichment cultures, a reliable strategy for
 isolation of complete operons and genes from microbial consortia
 JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
 MEDLINE 20575196
 PUBMED 11133432
 REFERENCE 2 (bases 1 to 4000)
 AUTHORS Entcheva, P., Liebl, W. and Streitz, W.R.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
 FEATURES
 source
 1..4000
 /organism="uncultured bacterium pCosFS1"

```

/db_xref="taxon:143795"
/clone="pCosFS1"
/note="unknown organism, cosmid clone derived from
environmental consortium"
complement(<1..1177)
/gene="biob"
complement(<1..1177)
/gene="biob"
/codon_start=1
/transl_table=11
/product="BioA-like protein"
/protein_id="AAG60558.1"
/db_xref="GI:12620098"
/translation="MSVTDSDLEPDQRIHPIPTYSMSHPDPCYPEAASVELQADG
BALQCVPLADSSVAEVELKMAALQWQARGRRRLTLRRHGIDFGAMSCDDP
NSMHLQGYLAHLFATAPQCGDAMDNDIAFAALLQHAQEVAAVLEEVQAG
AGMRITHTYLERVELCDRYOVLLIADDELATGPGRTGLFACHAHVVDILCLG
ALTGSYMTLSATLTRRHVAERTISNCAACGPHGPNPNCIACAVANASLSLLENRY
QOOVAITETOLKQULPLKMPKRVADVRLAIGVEVREPEVDMAQLODRCFVERGVM1"
gene
1265..2305
/gene="biob"
1265..2305
/gene="biob"
/codon_start=1
/transl_table=11
/product="BioB-like protein"
/protein_id="AAG60559.1"
/db_xref="GI:12620099"
/translation="MMADRIHWTVGLAOTLEPKLLEFEAQVYRHOPPROVQVS
TLSTKTGACPEDCXCPQSSRYKTGLSEERLMQVEYLESRRKAKANGSTRFCGAA
WNPHERDMPYLDQWQVGVKAMGECMTLGLDSQAEERLAEGLDYNHMLDTPPE
FYGNITTESYQERDITLQKVGAGIKVCSGIVGLGETVXXRAGLILVOLNLPPE
SVPIMLVVKGTPLADNDVDPEFIPTIIVARTMPSSTYRLSAGREONECQOAM
CFMAGNSIFYGCKLITTPNPEDKDLQFLKLGINPOOTTEHGDNQOQEVILARLM
QADTAEPYNAAL"
2305..3456
/gene="biob"
2305..3456
/gene="biob"
/note="7-keto-8-aminopelargonic acid synthetase; similar
to BioF"
/codon_start=1
/transl_table=11
/product="KAPA-synthetase"
/protein_id="AAG60556.1"
/db_xref="GI:12620096"
/translation="MSMOORIEQALEERYQSAAYHRLVTKGNGRYLQINDORYLNF
SGNDYIGLSQNNQVITAMQOQAOQYVSGSGSHYTGCTAQAQLEPQOLAOMLGPRS
LFTISGVANQALALAMQGEDRIILADRLSHASLLEAAHSPAQLRRFRHNPQMLAD
LIAKPCSGOTLAVTEGVFSMDSDSAPLAEVRLTRDAGAMLVADAHGAGVQEGCG
SCMOQCVRELLVTEFGAKFGLGAVALDESTAEYLQFARHLIYSTAMPQAQAL
OALHCIOQDELRTLRNTIOFRQGAOLSLTLESLETAIOLPLVGNQALDLAQ
HTRGQIMYSAIRPTVPPPGARLRTTLTAHQPODIDRLLEVLIDVSY"
3440..>4000
/gene="biob"
3440..>4000
/gene="biob"
/note="similar to BioC; probably reaction step prior to
synthesis of plimeloyl-CoA"
/codon_start=1
/transl_table=11
/product="biotin biosynthesis protein BioC"
/protein_id="AAG60557.1"
/db_xref="GI:12620097"
/translation="MTSAFEQVKNQAVASAFSRAAVGYDAAVLOHVGECRLGQOE
HOGGOLLADGCGTGYFSRRRELKQVTLADLAPARQOQADHYLLGIEVFP
LGDAAVDISFSSLYVQWCSLPRALAEIVRVRPGVILFSTLAEBSLHLELDAMQV
DGRHYNDFLPLQITACACAGYRHHLQ"
BASE COUNT 794 a 1164 c 1236 g 802 t 4 others
ORIGIN

```

```

Query Match 87.2%; Score 21.8; DB 2; Length 4000;
Best Local Similarity 92.0%; Pred. No. 2.9;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 gcattcgatcctcgtagt 25
||||| ||||| ||||| |||||
Db 1366 GCATTTCATCCGCTCAGTGCAG 1390

```

```

RESULT 17
ECOBIOB 128 bp DNA BCT 03-JAN-1995
LOCUS
DEFINITION Escherichia coli biotin (biob) gene, early terminator region.
ACCESSION M27731.1
VERSION M27731.1 GI:341755
KEYWORDS biob gene; biotin.
SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 128)
AUTHORS Nath,S.K.
TITLE Attenuation of transcription of biotin genes in Escherichia coli
JOURNAL Can. J. Microbiol. 34 (12), 1288-1296 (1988)
MEDLINE 89167942
FEATURES
source
1..128
location/Qualifiers
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
4..128
/gene="biob"

```

```

gene
-10..signal
4..9
/gene="biob"
20..>128
/gene="biob"
103..121
/gene="biob"
terminator
/note="early terminator"

```

```

BASE COUNT 31 a 28 c 34 g 35 t
ORIGIN

```

```

Query Match 84.0%; Score 21; DB 2; Length 128;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcattcgatcctcgtagt 21
||||| ||||| ||||| |||||
Db 59 GCATTTCATCCGCTCAGTGCAGT 79

```

```

RESULT 18
SMABIO 7215 bp DNA BCT 04-FEB-1999
LOCUS
DEFINITION S. marcescens biotin operon, complete sequence.
ACCESSION D17468
VERSION D17468.1 GI:402530
KEYWORDS 7,8-diamino-pelargonic acid aminotransferase;
7-keto-8-aminopelargonic acid synthetase; DAPA aminotransferase;
DPA synthetase; KAPA synthetase; biotin operon; biotin synthetase;
dehydrobiotin synthetase.

```

```

SOURCE
ORGANISM Serratia marcescens (strain:Str41) DNA.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Serratia.

```

```

REFERENCE 1 (bases 1 to 7215)
AUTHORS Sakurai,N.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-1993) to the DDBJ/EMBL/GenBank databases. Naoki
Sakurai, Tanabe Selyaku Co., Ltd., Res Lab of Applied Biochemistry,
2-50, Kawagishi-2-chome, Toda-shi, Saitama 335, Japan

```



```

gene
  /db_xref="taxon:549"
  1..1041
  /gene="D108"
  1..1041
  /gene="D108"
  /codon_start=1
  /transl_table=1
  /product="biotin synthetase"
  /protein_id="AAC4534.1"
  /db_xref="GI:1228114"
  /translation="MAHARWTLISQVTELPDKPLDLLFEAQTTHRQHFDPROYVST
  LKSTGACPECKYCPQSSRYKGLFETRLMEVQVDSARKAKNAGSTFPCGAAM
  KNPHERDMPYLEOMVQVKGAMGLEACMTLGLDETQAOQLASAGLDYNNHIDISPEF
  YGNITTTTYQRRLDTLDKVRDAGIKVCSGIVGIGFVVKRAGILLQLANIPRPES
  VPIINLVKGTPIADNDVDADFRTITAVARIWPTSEVRLSAGREOMNDQTOAMC
  FMAGANSIFVYCGCKLLTPNPEDKDVOLFRIKGLINPQOTDVMTGDNQOKLEQIFN
  ADTDFYNAAAL"

BASE COUNT      230 a      322 c      308 g      181 t

ORIGIN
Query Match      74.4%; Score 18.6; DB 3; Length 1041;
Best Local Similarity 84.0%; Pred. No. 1.1e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 gcaattcgatccctgcagtgacag 25
      ||| ||| ||| ||| ||| ||| |||
      99 GCACCTTCGACCTCCGCGAGGTTGAG 123

RESULT 20
LOCUS      AF165146      109723 bp      DNA      HTG      01-JUN-2000
DEFINITION Homo sapiens chromosome 8 map 8q12-8q13 clone CTA-397H3, WORKING
ACCESSION AF165146
VERSION AF165146.2 GI:8151945
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 109723)
REFERENCE Schudy A., Blechschmidt K., Menzel U., Pollay A., Reichwald K.,
AUTHORS Rump A., Schilhabel M.B., Taudien S., Wen G., Schlegelberger B.,
JOURNAL Siebert R. and Rosenthal A.
TITLE Chromosome 8 genomic sequence
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 109723)
JOURNAL Weber J., Schattervoy R. and Rosenthal A.
TITLE Direct Submission
AUTHORS Submitted (06-JUL-1999) Genome Analysis, Institute of Molecular
JOURNAL Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
TITLE 3 (bases 1 to 109723)
JOURNAL Genome Sequencing Center Jena.
COMMENT Direct Submission
Submitted (24-MAY-2000) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
On Jun 1, 2000 this sequence version replaced gi:5514637.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 61961: contig of 61961 bp in length
* 61962 62061: gap of unknown length
* 62062 91746: contig of 29885 bp in length
* 91747 91846: gap of unknown length
* 91847 109723: contig of 17877 bp in length.

```

```

FEATURES
  source
    Location/Qualifiers
    1..109723
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="8 map 8q12-8q13"
    /clone="CTA-397H3"
    1..61961
    /note="assembly-fragment
    clone-end:SP6
    vector-side:left"
    91647..109723
    /note="assembly-fragment
    clone-end:T7
    vector-side:right"

  misc_feature
    misc_feature
    91647..109723
    /note="assembly-fragment
    clone-end:T7
    vector-side:right"

BASE COUNT      31869 a      22757 c      23188 g      31709 t      200 others

ORIGIN
Query Match      74.4%; Score 18.6; DB 78; Length 109723;
Best Local Similarity 84.0%; Pred. No. 1.4e+02;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 gcaattcgatccctgcagtgacag 25
      ||| ||| ||| ||| ||| ||| |||
      Db 22635 GCATTTCCGCTCCTCAGAGTGCGAG 22659

RESULT 21
LOCUS      AC021450      176494 bp      DNA      HTG      04-APR-2000
DEFINITION Homo sapiens clone RP11-65E19, WORKING DRAFT SEQUENCE, 16 unordered
ACCESSION AC021450
VERSION AC021450.3 GI:7407964
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL 1 (bases 1 to 176494)
REFERENCE Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
AUTHORS Anderson S., Baldwin J., Barna N., Becker R., Beda F.,
JOURNAL Boguslavskiy L., Boukhgalter B., Brown A., Burkett G., Castle A.,
REFERENCE Choquel Y., Colangelo M., Collins S., Collymore A., Cooke P.,
AUTHORS Dearellano K., Dewar K., Domino M., Doyle M., Feneslor J.,
JOURNAL Ferreira P., Fitzhugh W., Forrest C., Gage D., Galagan J.,
REFERENCE Gardyna S., Grant G., Hagos B., Heatford A., Horton L.,
AUTHORS Howland J.C., Johnson R., Jones C., Kann L., Kartas A., Klein J.,
JOURNAL Landers T., Lehotzky J., Levine R., Lieu C., Liu G., Locke K.,
REFERENCE MacDonald P., Marquis N., McEwan P., McGuck A., McKernan K.,
AUTHORS McPeckers R., Meldrum J., Menus L., Morrow J., Naylor J.,
JOURNAL Norman C.H., O'Connor T., O'Donnell P., Olivari T.M., Peterson K.,
REFERENCE Pierre N., Pisanic C., Pollara V., Raymond C., Riley R., Rothman D.,
AUTHORS Roy A., Santos R., Severy P., Spencer B., Stange-Thomann N.,
JOURNAL Stojanovic N., Sudrmanthan A., Talamas J., Testafaye S., Theodore J.,
REFERENCE Tirrell A., Vassiliev H., Viel R., Vo A., Wu X., Wyman D., Ye W.,
AUTHORS Zimmer A. and Zody M.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 4, 2000 this sequence version replaced gi:6814993.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

```

```

--
-- Project Information
Center project name: L5159
Center clone name: 65_E.19

-- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167673 bases at least Q40
Consensus quality: 171804 bases at least Q30
Consensus quality: 173582 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 174994; sum-of-coverage
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.8 in Q20 bases; sum-of-coverage
--
--
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*
1      949: contig of 949 bp in length
*
*      950 1049: gap of 100 bp
*
*      1050 2544: contig of 1495 bp in length
*
*      2545 2644: gap of 100 bp
*
*      2645 5857: contig of 3213 bp in length
*
*      5858 5957: gap of 100 bp
*
*      5958 9641: contig of 3684 bp in length
*
*      9642 9741: gap of 100 bp
*
*      9742 14928: contig of 5187 bp in length
*
*      14929 15028: gap of 100 bp
*
*      15029 20347: contig of 5319 bp in length
*
*      20348 20447: gap of 100 bp
*
*      20448 27447: contig of 7000 bp in length
*
*      27448 27547: gap of 100 bp
*
*      27548 39249: contig of 11702 bp in length
*
*      39250 39349: gap of 100 bp
*
*      39350 46799: contig of 7450 bp in length
*
*      46800 46899: gap of 100 bp
*
*      46900 56544: contig of 9645 bp in length
*
*      56545 56644: gap of 100 bp
*
*      56645 65562: contig of 8918 bp in length
*
*      65563 65662: gap of 100 bp
*
*      65663 81928: contig of 16266 bp in length
*
*      81929 82028: gap of 100 bp
*
*      82029 97874: contig of 15846 bp in length
*
*      97875 97974: gap of 100 bp
*
*      97975 118702: contig of 20728 bp in length
*
*      118703 118802: gap of 100 bp
*
*      118803 142786: contig of 23984 bp in length
*
*      142787 142886: gap of 100 bp
*
*      142887 176494: contig of 33608 bp in length.
*
FEATURES
source
1. .176494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-65E19"
/clone_lib="RPC1-11 Human Male BAC"
misc_feature
1. .949
/note="assembly_fragment
clone_end:r7
vector_side:left"
misc_feature
1050. .2544
/note="assembly_fragment"
2645. .5857
/note="assembly_fragment"
5958. .9641
/note="assembly_fragment
clone_end:SP6
vector_side:left"
misc_feature
9742. .14928

```

```
misc_feature /note="assembly_fragment" 15029. . 20347
misc_feature /note="assembly_fragment" 15029. . 20347
misc_feature /note="assembly_fragment" 20448. . 27447
misc_feature /note="assembly_fragment" 27548. . 39249
misc_feature /note="assembly_fragment" 39350. . 46799
misc_feature /note="assembly_fragment" 46900. . 56544
misc_feature /note="assembly_fragment" 56645. . 65562
misc_feature /note="assembly_fragment" 65663. . 81928
misc_feature /note="assembly_fragment" 82029. . 97874
misc_feature /note="assembly_fragment" 97975. . 118702
misc_feature /note="assembly_fragment" 118803. . 142786
misc_feature /note="assembly_fragment" 142887. . 176494
misc_feature /note="assembly_fragment" 176494. . 313595
misc_feature /note="assembly_fragment" 313595. . 553633
misc_feature /note="assembly_fragment" 553633. . 1503000
```

Query Match:	74.48%	Score 18.6:	DB 66,	Length 176694,
Best Local Similarity:	84.0%;	Pred. No. 1.4e+02;		
Matches	21;	Conservative	0;	Mismatches 4;
			Indels	0;
			Gaps	0;
Qy	1	gcatttcgacatccctgcagtcgacg	25	
Db	89764	GCACCTTCACATCCGCGCCACGATGCAC	89788	

RESULT	22
LOCUS	AC067950
DEFINITION	AC067950 186733 bp DNA HTG 03-APR-2001
ACCESSION	Homo sapiens chromosome YUK clone RP11-375B23, WORKING DRAFT
VERSION	AC067950
KEYWORDS	AC067950.4 GI:13518257
SOURCE	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEPIN.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
JOURNAL	1 (bases 1 to 186733)
REFERENCE	Waterston, R.H.
AUTHORS	The sequence of Homo sapiens clone
TITLE	Unpublished
JOURNAL	2 (bases 1 to 186733)
REFERENCE	Waterston, R.H.
AUTHORS	Direct Submission
TITLE	Submitted (27-APR-2000). Genome Sequencing Center, Washington
JOURNAL	University School of Medicine, 4444 Forest Park Parkway, St. Louis,
COMMENT	MO 63108, USA
	On Apr 3, 2001 this sequence version replaced gi:13357540.

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H.NH0375B23
----- Summary Statistics -----
Sequencing vector: M13; 68%
Sequencing vector: plasmid; 27%
Chemistry: Dye-Primer ET; 68% of reads
Chemistry: Dye-terminator Big Dye; 27% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 185851 bases at least Q40

```

Consensus quality: 186234 bases at least Q30
 Consensus quality: 186596 bases at least Q20
 Insert size: 175000; agarose-fp
 Insert size: 186733; sum-of-ctrls
 Quality coverage: 7.49 in Q20 bases; agarose-fp
 Quality coverage: 7.02 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 1 186733: contig of 186733 bp in length.

Location/Qualifiers

source
 1. 186733
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="UNK"
 /clone="RP11-375B23"
 /note="assembly_name:Contig4
 clone_end:SP6
 vector_side:left"

misc-feature
 1. 186733
 /note="assembly_name:Contig4
 clone_end:SP6
 vector_side:left"

BASE COUNT 52694 a 40148 c 42452 g 51438 t 1 others

ORIGIN

Query Match 74.4%; Score 18.6; DB 73; Length 186733;
 Best Local Similarity 84.0%; Pred. No. 1.4e+02;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gccattgcattcctcgtcagtcagc 25

Db 161876 GCATTGGACCTGCTGCTGACG 161900

RESULT 23

AC017911

LOCUS AC017911 118100 bp DNA 09-DEC-1999

DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered

pieces.

AC017911

VERSION AC017911.1 GI:6553279

KEYWORDS

HTG: HTGS_PHASE2.

SOURCE

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 118100)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 118100

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="49A-49B"

/map="BACR37P06 (D594)"

/clone="BACR37P06 (D594)"

/clone="BACR37P06 (D594)"

/clone="BACR37P06 (D594)"

/clone="BACR37P06 (D594)"

/clone="BACR37P06 (D594)"

/clone="BACR37P06 (D594)"

/clone="BACR37P06 (D594)"

/clone="BACR37P06 (D594)"

/clone="BACR37P06 (D594)"

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gccattgcattcctcgtcagc 20

Db 63425 GCATTGGACCTGCTGCTGACG 63444

RESULT 24

AC007476/c

LOCUS

DEFINITION

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

AC007476

Query Match 73.6%; Score 18.4; DB 4; Length 171705;
 Best Local Similarity 95.0%; Pred. No. 1.8e+02;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 gatttcgattctgctcagg 20
 |||||
 Db 106124 GCATTTCGATCTCGTAAG 106105

RESULT 25
 AEO03822/c
 LOCUS AEO03822 266133 bp DNA INV 04-OCT-2000
 DEFINITION Drosophila melanogaster genomic scaffold 142000013386047 section 21
 of 52, complete sequence.
 ACCESSION AEO03822 AEO02787
 VERSION AEO03822.2 GI:10727606
 KEYWORDS HTG.
 SOURCE fruit fly,
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 I (bases 1 to 266133)
 REFERENCE
 AUTHORS Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
 Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
 George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
 Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
 Brandon,R.C., Rogers,Y.H., Blazer,R.G., Champ,M., Pfeiffer,B.D.,
 Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gaber
 Milos,G.L., Abri,J.F., Agbayani,A., An,H.J.,
 Andrews-Flannkoch,C., Baldwin,D., Bailew,R.M., Basu,A.,
 Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
 Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
 Bouchan,M.R., Bouck,J.J., Brokstein,P., Brotlier,P., Burtis,K.C.,
 Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
 Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
 Pablo,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
 Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
 Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
 Fleischmann,W., Foster,C., Gabriellian,A.E., Garg,N.S.,
 Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
 Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
 Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
 Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
 Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
 Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
 McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
 Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
 Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
 Nusskern,D.R., Paclet,J.M., Palazotto,M., Pittman,G.S., Pan,S.,
 Pollard,D., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
 Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
 Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
 Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
 Wooley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
 Zaveri,J.C., Zhu,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
 The genome sequence of Drosophila melanogaster
 Science 287 (5461), 2185-2195 (2000)
 20196006
 2 (bases 1 to 266133)
 REFERENCE
 AUTHORS Adams,M.D., Celisner,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.
 TITLE Direct Submission
 JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 COMMENT On Oct 9, 2000 this sequence version replaced gi:7303437.
 FEATURES
 Location/Qualifiers

source
 1..266133
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="2R"
 join(392..425,1007..1155,5710..6105)
 /gene="Cam"
 /product="CT41657"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 <392..>15192
 /gene="Cam"
 /note="CG8472"
 /map="48P5-48P5"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(405..425,1007..1155,5710..5884,10160..10402,
 14732..15192)
 /gene="Cam"
 /product="CT41655"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(596..619,1004..1155,5710..5884,10160..10402,
 14732..14876)
 /gene="Cam"
 /product="CT41659"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(1153..1155,5710..5922)
 /gene="Cam"
 /note="Cam gene product [alt 1]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58544.1"
 /db_xref="GI:7303488"
 /translation="MADOLTEQIAEFKFAESLPDKDGGTITTKELGTVRSIGONP
 TAEIADMNEVDADNGITDPEFLTMARKMDTDEEELREAFRFDDGNGFIS
 join(1153..1155,5710..5884,10160..10402,14732..14760)
 /gene="Cam"
 /note="Cam gene product [alt 4]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58542.1"
 /db_xref="GI:7303486"
 /translation="MADOLTEQIAEFKFAESLPDKDGGTITTKELGTVRSIGONP
 TAEIADMNEVDADNGITDPEFLTMARKMDTDEEELREAFRFDDGNGFIS
 join(1153..1155,5710..5884,10160..10402,14732..14760)
 /gene="Cam"
 /note="Cam gene product [alt 2]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58541.1"
 /db_xref="GI:7303485"
 /translation="MADOLTEQIAEFKFAESLPDKDGGTITTKELGTVRSIGONP
 TAEIADMNEVDADNGITDPEFLTMARKMDTDEEELREAFRFDDGNGFIS
 AAEIHWMTNIGEKLTDEEVDMEIADIDGGVNYEEFYMTSK"
 join(5527..5642,5710..5884,10160..10402,14732..15192)
 /gene="Cam"
 /product="CT24787"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(5640..5642,5710..5884,10160..10402,14732..14760)
 /gene="Cam"
 /note="Cam gene product [alt 3]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58543.1"
 /db_xref="GI:7303487"
 /translation="MADOLTEQIAEFKFAESLPDKDGGTITTKELGTVRSIGONP
 TAEIADMNEVDADNGITDPEFLTMARKMDTDEEELREAFRFDDGNGFIS
 AAEIHWMTNIGEKLTDEEVDMEIADIDGGVNYEEFYMTSK"

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1004: contig of 1004 bp in length
 * 1005 1104: gap of 100 bp
 * 1105 7217: contig of 6113 bp in length
 * 7218 7317: gap of 100 bp
 * 7318 18994: contig of 11677 bp in length
 * 18995 19094: gap of 100 bp
 * 19095 35888: contig of 16794 bp in length
 * 35889 35988: gap of 100 bp
 * 35989 52052: contig of 16064 bp in length
 * 52053 52152: gap of 100 bp
 * 52153 69573: contig of 17421 bp in length
 * 69574 69673: gap of 100 bp
 * 69674 93465: contig of 23792 bp in length
 * 93466 93565: gap of 100 bp
 * 93566 116170: contig of 22605 bp in length
 * 116171 116270: gap of 100 bp
 * 116271 154270: contig of 38000 bp in length.

FEATURES

source
 1. 154270
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /map="4"
 /clone="RP11-147K21"
 /clone_1lb="RPCT-11 Human Male BAC"
 1. 1004
 /note="assembly-fragment"
 1105. 7217
 /note="assembly-fragment"
 clone_end:T7
 vector_side:left"
 7318. 18994
 /note="assembly-fragment"
 clone_end:SP6
 vector_side:right"
 19095. 35888
 /note="assembly-fragment"
 35989. 52052
 /note="assembly-fragment"
 52153. 69573
 /note="assembly-fragment"
 69674. 93465
 /note="assembly-fragment"
 93566. 116170
 /note="assembly-fragment"
 116271. 154270
 /note="assembly-fragment"

BASE COUNT 52075 a 27858 c 27285 g 46252 t 800 others
 ORIGIN

Query Match 72.8%: Score 18.2; DB 68; Length 154270;
 Best Local Similarity 87.0%: Pred. No. 2.2e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cattgcatcctgcagtgca 24
 ||||| ||||| ||||| |||||
 Db 39854 CATTTGGCTCCTCGTCAATGTCA 39832

RESULT 27
 LOCUS AK024633 1315 bp mRNA PRI 29-SEP-2000
 DEFINITION Homo sapiens cDNA: FLT20980 fls, clone ADSU01986.
 ACCESSION AK024633
 VERSION AK024633.1 GI:10436953
 KEYWORDS Oligo capping; fls (full insert sequence).
 SOURCE Homo sapiens adipose tissue cDNA to mRNA, clone_1lb:ad
 clone:ADSU01986.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (sites)

Kawakami,T., Noguchi,S., Itoh,T., Shigetani,K., Senba,T.,
 Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,A.,
 Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hirao,M.,
 Ohmori,Y., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T.,
 Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
 NEDO human cDNA sequencing project
 unpublished (2000)

JOURNAL

2 (bases 1 to 1315)

Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
 Shibahara,T., Tanaka,T. and Nakamura,Y.
 Direct Submission
 Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
 Sugano, Institute of Medical Science, University of Tokyo,
 Laboratory of Genome Structure Analysis, Human Genome Center;
 Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
 (E-mail:cdna@elms.u-tokyo.ac.jp, Tel:81-3-5449-5286,
 Fax:81-3-5449-5416)

COMMENT

NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan: cDNA full insert
 sequencing: Research Association for Biotechnology; cDNA library
 construction, 5'-3' end one pass sequencing: Department of
 Virology and Human Genome Center, Institute of Medical Science,
 University of Tokyo (partly supported by Science and Technology
 Agency).

FEATURES

source
 1. 1315
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="ADSU01986"
 /clone_1lb="ad"
 /tissue_type="adipose tissue"
 /note="Cloning vector pME18SFL1"
 283. 1257
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB14939.1"
 /db_xref="GI:10436954"
 /translation="MESLFPYGPPTPOTLAGISGALLQYQYQYQOMLOESLQKQKQ
 QOEQOQKPVQAKTSKESKESDOPQNSNDASKEDEKSDSTESTREPOLSKESKADSDY
 VYFEVYKFCRCOMFTDEDAVNHQSKSCYCGOPLIDQETVLRYVYKCYCLAG
 DVAISGENALSOHLOSLHKEKTTIOAARNKKEVRLHPVSGPNNNTSTGSSAS
 SNNTYPHLSCEFSKMSWPIILFOASARRASPPSSPSISLPYTSYICSGSVQVTSLS
 PRESQSDSDSELSKRLDLDNLSLEVRKKRPAAGLDGPNSTRMDPSY"

CDS

BASE COUNT 409 a 326 c 266 g 314 t
 ORIGIN

Query Match 70.4%: Score 17.6; DB 89; Length 1315;
 Best Local Similarity 83.3%: Pred. No. 3.7e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 cattgcatcctgcagtgca 25
 ||||| ||||| ||||| |||||
 Db 115 CATGTGATCGATTCAGTTCAGTTCAG 138

RESULT 28
 AF298117/c 10728 bp DNA PRI 26-SEP-2000
 LOCUS AF298117
 DEFINITION Homo sapiens homeobox protein OTX2 gene, complete cds.
 ACCESSION AF298117
 VERSION AF298117.1 GI:10304407
 KEYWORDS human.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 10728)
 Perrault,I., Rozet,J.-M., Gerber,S., Munnich,A. and Kaplan,J.

Best Local Similarity 83.3%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 gattcgtacccgtcagtgca 24
||||| | | | | | | | | |
Db 51662 GCATTCCATCTCTGTCAGACCA 51685

RESULT 31

AC005646 86677 bp DNA HTG 10-DEC-1999
LOCUS Drosophila melanogaster chromosome 2 clone DS00968 (D404) map
DEFINITION 51B1-51B4 strain y; cn bw sp, *** SEQUENCING IN PROGRESS ***, 3
unordered pieces.

AC005646
VERSION AC005646.6 GI:6554236
KEYWORDS HTG; HTGS; PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

REFERENCE
AUTHORS Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 86677)

REFERENCE
AUTHORS Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, R.A., Harris, M.L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, M.L.,
Hinkle, L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Katta, K.,
Keane, L., Lee, B., Lewis, S., Li, P., Ling, H., Mostreli, A.R.,
Mostreli, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B.,
Richards, S., Sethi, H., Svirska, R.R., Wan, K.H., Webster, D.,
Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 86677)

REFERENCE
AUTHORS Celniker, S.E., George, R.A., Galle, R., Svirska, R.R., Hoskins, R.A.,
Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G., Chavez, C.,
Chew, R.A., Doyle, C.M., Farfan, D.E., Flanagan, J., Houston, K.A.,
Hummasti, S.R., Katta, K., Kearney, L., Kim, S.H., Lee, B.,
Lomocan, M.A., Mak, J., Mazda, P., Mok, M.S., Mostreli, A.R.,
Mostreli, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Punch, E.,
Salt, E., Twomey, B., Wan, K.H., Whitelaw, K.R., Yee, A., Zhang, R.,
Zieran, L.L. and Kimmel, B.E.
Direct Submission
Submitted (10-SEP-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Dec 10, 1999 this sequence version replaced gi:55656719.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to hdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases. Pl library location:
8-11.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 296: contig of 296 bp in length
* 297 376: gap of unknown length
* 377 85947: contig of 85571 bp in length
* 85948 86027: gap of unknown length
* 86028 86677: contig of 650 bp in length.

FEATURES

source
1. 86677
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2"
/map="51B1-51B4"
/clone="DS00968 (D404)"

/clone.lib="Pl library, partial Sau3A in pNS582tel14Ad10"
BASE COUNT 24015 a 18183 c 18237 g 26082 t 160 others
ORIGIN

Query Match 70.4%; Score 17.6; DB 60; Length 86677;
Best Local Similarity 83.3%; Pred. No. 4.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 cattcgtacccgtcagtgca 25
|| | | | | | | | | | | | | | |
Db 60896 CAATTGCATCTCTGTCAGACGAG 60919

RESULT 32

LMFLCHR15_6
WPCOMMENT
Sequence split into 8 fragments LOCUS LMFLCHR15 Accession AL160371

Fragment Name	Begin	End
LMFLCHR15_0	1	110000
LMFLCHR15_1	100001	210000
LMFLCHR15_2	200001	310000
LMFLCHR15_3	300001	410000
LMFLCHR15_4	400001	510000
LMFLCHR15_5	500001	610000
LMFLCHR15_6	600001	710000
LMFLCHR15_7	700001	720704

Continuation (7 of 8) of LMFLCHR15 from base 600001 (AL160371) Leishmania major chromo

Query Match 70.4%; Score 17.6; DB 84; Length 110000;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 gattcgtacccgtcagtgca 24
||||| | | | | | | | | |
Db 83492 GCATTCCATCTCTGTCAGACCA 83515

RESULT 33

HSUJ335E1 143805 bp DNA PRI 25-MAY-2000
LOCUS Human DNA sequence from clone RP3-335E1 on chromosome 6q16.2-21
DEFINITION Contains STSs, GSSs and a Cpg Island, complete sequence.
ACCESSION AL109920
VERSION AL109920.15 GI:7406500
KEYWORDS HTG: Cpg Island.
SOURCE human.

REFERENCE
AUTHORS Homo sapiens
1 (bases 1 to 143805)
Frankland, J.
Direct Submission
Submitted (28-APR-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Apr 3, 2000 this sequence version replaced gi:7378738.

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unSURE'
feature key.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMBASE; Information

on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 RP3-335E1 is from the library RP3-3 constructed at the Roswell
 Park Cancer Institute by the group of Pieter de Jong. For further
 details see <http://bacpac.med.buffalo.edu/>
 VECTOR: pCYPAC2

This sequence is the entire insert of clone RP3-335E1.

FEATURES

Source

Location/Qualifiers
 1..143805
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="q16.2-21"
 /clone="RP3-335E1"
 /clone_lib="RP3-3"
 9..54
 /note="L2 repeat: matches 2705..2750 of consensus"
 546..638
 /note="LTR3 repeat: matches 256..355 of consensus"
 693..849
 /note="MER69A repeat: matches 22..175 of consensus"
 849..963
 /note="LTR3 repeat: matches 403..518 of consensus"
 2404..2638
 /note="L2 repeat: matches 2491..2743 of consensus"
 complement(2642..3106)
 /note="match: GSS: Em:AQ629085"
 complement(2668..3066)
 /note="match: GSS: Em:B85272"
 2947..4161
 /note="LIM47 repeat: matches 5095..6282 of consensus"
 4208..4526
 /note="AlusX repeat: matches 1..312 of consensus"
 4528..4625
 /note="AlusX repeat: matches 1..312 of consensus"
 4776..5058
 /note="49 copies 2 mer aa 65% conserved"
 5549..5858
 /note="AlusX repeat: matches 17..301 of consensus"
 5897..5996
 /note="AlusX repeat: matches 1..300 of consensus"
 9232..9688
 /note="MIR repeat: matches 76..182 of consensus"
 9698..10507
 /note="LIMB2 repeat: matches 5769..6168 of consensus"
 10676..10709
 /note="LIP10 repeat: matches 5295..6164 of consensus"
 10805..10844
 /note="17 copies 2 mer aa 82% conserved"
 complement(1164..11305)
 /note="match: GSS: Em:AQ070712"
 11450..11804
 /note="match: GSS: Em:AQ253061"
 13851..14130
 /note="AlusX repeat: matches 1..312 of consensus"
 14304..14351
 /note="LIM3d repeat: matches 251..300 of consensus"
 15500..16053
 /note="LIM2 repeat: matches 697..1258 of consensus"
 16056..16439
 /note="match: GSS: Em:AQ123207"
 16105..16169
 /note="LIM2 repeat: matches 1252..1315 of consensus"
 16143..16306
 /note="LIM2 repeat: matches 1824..2364 of consensus"
 16257..16436
 /note="match: GSS: Em:AQ508275 Em:AQ750035 Em:B85162"
 16505..17689
 /note="LIM1 repeat: matches 1908..3130 of consensus"

repeat_region 17686..18668
 /note="LIM6 repeat: matches 3638..4613 of consensus"
 repeat_region 18669..18963
 /note="Aluo repeat: matches 1..290 of consensus"
 repeat_region 18964..19047
 /note="LIM6 repeat: matches 4613..4695 of consensus"
 repeat_region 19048..19378
 /note="Aluo repeat: matches 1..305 of consensus"
 repeat_region 19379..19801
 /note="LIM6 repeat: matches 4695..5139 of consensus"
 repeat_region 19780..20811
 /note="LIM6 repeat: matches 5285..6297 of consensus"
 repeat_region 20856..21173
 /note="Aluy repeat: matches 1..311 of consensus"
 repeat_region 21177..21256
 /note="MTTIE repeat: matches 11..81 of consensus"
 repeat_region 21450..21616
 /note="MTTIE repeat: matches 286..460 of consensus"
 repeat_region 21682..21815
 /note="LTR3 repeat: matches 374..521 of consensus"
 repeat_region 21914..22119
 /note="LTR3 repeat: matches 5..221 of consensus"
 repeat_region 22394..22698
 /note="Aluy repeat: matches 1..304 of consensus"
 repeat_region 22698..22971
 /note="AlusX repeat: matches 7..281 of consensus"
 repeat_region 22984..23284
 /note="Aluy repeat: matches 1..298 of consensus"
 repeat_region 23528..24208
 /note="LIM5 repeat: matches 4919..5626 of consensus"
 repeat_region 24212..24350
 /note="MTTIA2 repeat: matches 4..143 of consensus"
 repeat_region 24346..24422
 /note="MTTIA2 repeat: matches 294..374 of consensus"
 repeat_region 24423..24695
 /note="LIM5 repeat: matches 5622..5921 of consensus"
 repeat_region 24697..25040
 /note="MER1A repeat: matches 1..347 of consensus"
 repeat_region 26627..26852
 /note="MIR repeat: matches 5..262 of consensus"
 repeat_region 27852..28161
 /note="L2 repeat: matches 2074..2398 of consensus"
 repeat_region 29319..29456
 /note="FRAM repeat: matches 1..151 of consensus"
 repeat_region 30147..30209
 /note="LIM3 repeat: matches 6602..6668 of consensus"
 repeat_region 30738..31054
 /note="Alub repeat: matches 1..306 of consensus"
 repeat_region 31861..32423
 /note="match: STS: Em:AA134931"
 repeat_region 31989..32126
 /note="match: STS: Em:G44108"
 repeat_region 33086..33248
 /note="AlusX repeat: matches 127..289 of consensus"
 repeat_region 33329..33490
 /note="Alub repeat: matches 136..296 of consensus"
 repeat_region 33675..33782
 /note="L2 repeat: matches 2096..2209 of consensus"
 repeat_region 34176..34821
 /note="L2 repeat: matches 2056..2710 of consensus"
 repeat_region 34972..35128
 /note="Alub repeat: matches 132..288 of consensus"
 repeat_region 35810..35950
 /note="MER46C repeat: matches 59..196 of consensus"
 repeat_region 36166..36486
 /note="Alub repeat: matches 1..312 of consensus"
 repeat_region 38407..38437
 /note="L2 repeat: matches 2690..2720 of consensus"
 repeat_region 39159..39454
 /note="AlusX repeat: matches 1..296 of consensus"
 repeat_region 40502..40795
 /note="Aluo repeat: matches 1..294 of consensus"
 repeat_region 41205..41505

```

repeat_region /note="AluY repeat: matches 1. .308 of consensus"
42583. .42890
/note="AluSg repeat: matches 1. .310 of consensus"
repeat_region 43381. .43376
/note="L1ME repeat: matches 5304. .5395 of consensus"
repeat_region 43377. .43652
/note="AluY repeat: matches 1. .310 of consensus"
repeat_region 43653. .44046
/note="L1ME repeat: matches 5395. .5813 of consensus"
repeat_region 45125. .45371
/note="L2 repeat: matches 2373. .2626 of consensus"
repeat_region 46153. .46351
/note="MSTC repeat: matches 1. .183 of consensus"
repeat_region 46456. .46593
/note="MSTC repeat: matches 253. .394 of consensus"
misc.feature 46601. .46944
/note="match: GSS: Em:AQ097981"
repeat_region 47206. .47513
/note="AluSx repeat: matches 1. .300 of consensus"
repeat_region 48740. .48838
/note="H3 repeat: matches 2. .100 of consensus"
repeat_region 50464. .50617

```

```

Query Match 70.4%: Score 17.6; DB 93; Length 143805;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```
QY 2 catctgacatcctgcagctgacg 25
||||| ||||| ||||| |||
Db 90879 CATTTGAGCCTCATTCAGCTTCAG 90902
```

```

RESULT 34
AC090711 153063 bp DNA PRI 09-MAR-2001
LOCUS AC090711 153063 bases, complete sequence.
DEFINITION AC090711
ACCESSION AC090711
VERSION AC090711.1 GI:13259420
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 153063)
AUTHORS Wen,G., Polley,A., Baumgart,C., Schillhabl,M., Rosenthal,A. and
Platzner,M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 153063)
AUTHORS Wen,G., Rosenthal,A., Schattevoy,R., Baumgart,C. and Platzner,M.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2001) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
1. 153063
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8q21.1"
/clone="RP11-89H1"
BASE COUNT 47757 a 27446 c 28431 g 49429 t
ORIGIN

```

```

Query Match 70.4%: Score 17.6; DB 88; Length 153063;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```
QY 2 catctgacatcctgcagctgacg 25
||||| ||||| ||||| |||
Db 121526 CATGTGATCTCTATTTCAGTTGCAG 121549
```

```

RESULT 35
AC023079/c

```

```

LOCUS AC023079 157254 bp DNA HTG 07-JUL-2000
DEFINITION Homo sapiens chromosome 14 clone RP11-59P19 map 14, WORKING DRAFT
SEQUENCE, 26 unordered pieces.
AC023079
AC023079.3 GI:8954144
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 157254)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 14, clone RP11-59P19
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 157254)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F.,
Boguslavsky,L., Bouckghalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fensholt,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Healdorf,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczeky,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Meldrum,J., Menus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rotman,D.,
Roy,A., Santos,R., Severny,P., Spencer,P., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (08-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 7, 2000 this sequence version replaced g1:7139783.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L6308
Center Clone name: 39_P_19
----- Summary Statistics
Sequencing vector: M13; M77815, 100% of reads
Sequencing vector: Plasmid; n/a; %0.1% of reads
0.4953998584571830chemistry: Dye-terminator Big Dye; 100% of
reads
Assembly program: Phrap; version 0.960731
Consensus quality: 142719 bases at least Q40
Consensus quality: 150008 bases at least Q30
Consensus quality: 152903 bases at least Q20
Insert size: 109000; agarose-ff
Insert size: 154754; sum-of-contris
Quality coverage: 6.1 in Q20 bases; agarose-ff
Quality C.
NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1263: contig of 1263 bp in length
* 1264 1363: gap of 100 bp
* 1364 2624: contig of 1261 bp in length
* 2625 2724: gap of 100 bp
* 2725 4379: contig of 1655 bp in length

```

```

* 4380 4479: gap of 100 bp
* 4480 6168: contig of 1689 bp in length
* 6169 6268: gap of 100 bp
* 6269 7502: contig of 1234 bp in length
* 7503 7602: gap of 100 bp
* 7603 9227: contig of 1625 bp in length
* 9228 9327: gap of 100 bp
* 9328 11272: contig of 1945 bp in length
* 11273 11372: gap of 100 bp
* 11373 14027: contig of 2655 bp in length
* 14028 14127: gap of 100 bp
* 14128 16796: contig of 2669 bp in length
* 16797 16896: gap of 100 bp
* 16897 19661: contig of 2765 bp in length
* 19662 19761: gap of 100 bp
* 19762 21731: contig of 1970 bp in length
* 21732 21831: gap of 100 bp
* 21832 25103: contig of 3272 bp in length
* 25104 25203: gap of 100 bp
* 25204 27943: contig of 2740 bp in length
* 27944 28043: gap of 100 bp
* 28044 32560: contig of 4517 bp in length
* 32561 32660: gap of 100 bp
* 32661 39551: contig of 6891 bp in length
* 39552 39651: gap of 100 bp
* 39652 44837: contig of 5186 bp in length
* 44838 44937: gap of 100 bp
* 44938 52684: contig of 7747 bp in length
* 52685 52784: gap of 100 bp
* 52785 60087: contig of 7303 bp in length
* 60088 60187: gap of 100 bp
* 60188 68029: contig of 7842 bp in length
* 68030 68129: gap of 100 bp
* 68130 76334: contig of 8205 bp in length
* 76335 76434: gap of 100 bp
* 76435 88231: contig of 11797 bp in length
* 88232 88331: gap of 100 bp
* 88332 99697: contig of 11366 bp in length
* 99698 99797: gap of 100 bp
* 99798 112971: contig of 13174 bp in length
* 112972 113071: gap of 100 bp
* 113072 125118: contig of 12047 bp in length
* 125119 125218: gap of 100 bp
* 125219 142067: contig of 16849 bp in length
* 142068 142167: gap of 100 bp
* 142168 157254: contig of 15087 bp in length.
*
* Location/Qualifiers
*   1..157254
*     organism="Homo sapiens"
*     db_xref="taxon:9606"
*     chromosome="14"
*     map="14"
*     clone.lib="RP11-59P19"
*     1..1263
*       misc_feature
*         /note="assembly-fragment"
*       1364..2624
*         misc_feature
*           /note="assembly-fragment"
*         2725..4379
*           misc_feature
*             /note="assembly-fragment"
*           4480..6168
*             misc_feature
*               /note="assembly-fragment"
*             6269..7502
*               misc_feature
*                 /note="assembly-fragment"
*               7603..9227
*                 misc_feature
*                   /note="assembly-fragment"
*                 9328..11272
*                   misc_feature
*                     /note="assembly-fragment"
*                     11373..14027
*                       misc_feature
*                         /note="assembly-fragment"
*                         14128..16796
*                           misc_feature
*                             /note="assembly-fragment"
*                             16897..19661

```

```

misc_feature /note="assembly-fragment"
19762..21731
/note="assembly-fragment"
21832..25103
/note="assembly-fragment"
25204..27943
/note="assembly-fragment
clone_end:r7
vector_side:left"
28044..32560
/note="assembly-fragment"
32661..39551
/note="assembly-fragment"
39652..44837
/note="assembly-fragment"
44938..52684
/note="assembly-fragment"
52785..60087
/note="assembly-fragment"
60188..68029
/note="assembly-fragment
clone_end:sp6
vector_side:left"
68130..76334
/note="assembly-fragment"
76435..88231
/note="assembly-fragment"
88332..99697
/note="assembly-fragment"
99798..112971
/note="assembly-fragment"
113072..125118
/note="assembly-fragment"
125219..142067
/note="assembly-fragment"
142168..157254
/note="assembly-fragment"
BASE COUNT 44228 a 32822 c 33152 g 44549 t 2503 others
ORIGIN
Query Match 70.4%; Score 17.6; DB 67; Length 157254;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 2 cattcgatcctcgtcaggtcag 25
||||| | ||||| || |||||
Db 125667 CATTGGGTCCTCGCAGCTGCAG 125644

```

```

RESULT 36
AC007580
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AC007580 160903 bp DNA INV 28-FEB-2001
Drosophila melanogaster, chromosome 2R, region 50F-51C, BAC clone
AC007580
AC007580.5 GI:13162480
HTG.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 160903)
Celinker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gocayne,J.D., Amaratilake,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busen,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorset,Y., Doupl,E., Doyle,C., Drenek,D., Farfan,D.,
Farrera,S., Frise,E., Galle,R.F., Gary,N.S., George,R.A.,
Gonzalez,M., Houch,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
Ileguam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,

```

Pacled, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pitman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svitskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M., and Venter, J.C.
Sequencing of Drosophila chromosome 2R, region 50F-51C
Unpublished
2 (bases 1 to 160903)

Celniker, S.E., Abmayyan, A., Arcaina, T.T., Baxter, E., Blazek, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomolan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacled, J.M., Park, S., Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Snir, E., Svitskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

Direct Submission
Submitted (18-MAY-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 28, 2001 this sequence version replaced gi:5670594.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES

source
1. 160903
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="50F-51C"
/clone="BACR07P02 (D623)"
/clone_1lb="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"
BASE COUNT 45488 a 33530 c 34013 g 47872 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 4; Length 160903;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 cattcgatcctgcgtcagtcag 25
||| ||||| ||||| |||||
Db 13960 CAATGGATCTGCTGACGACGAG 13983

RESULT 37
AC008306 169689 bp DNA INV 21-MAR-2001
LOCUS Drosophila melanogaster, chromosome 2R, region 51B-51B, BAC clone
DEFINITION BACR27M04, complete sequence.
AC008306
VERSION AC008306.3 GI:13399345
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

REFERENCE
AUTHORS
Celniker, S.E., Adams, M.D., Kronmiller, B., Tyler, D., Wan, K.H., Holt, R.A., Evans, C.A., Gocayne, J.D., Amand, L.S., Brandon, R.C., Rogers, Y., An, H., Baldwin, D., Bonzon, J., Beeson, K.Y., Busam, D.A., Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D.,

Perliera, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ileguam, C., Jallali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacled, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pitman, G.S., Puri, V., Richards, S., Scheeler, F., Stapleton, M., Strong, R., Svitskas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome 2R, region 51B-51B
Unpublished
2 (bases 1 to 169689)

Celniker, S.E., Abmayyan, A., Arcaina, T.T., Baxter, E., Blazek, R.G., Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomolan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacled, J.M., Park, S., Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Snir, E., Svitskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

Direct Submission
Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 21, 2001 this sequence version replaced gi:5738288.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES

source
1. 169689
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="51B-51B"
/clone="BACR27M04 (D708)"
/clone_1lb="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcoRI in pBAC3.6)"
BASE COUNT 48346 a 35214 c 34999 g 51130 t
ORIGIN

Query Match 70.4%; Score 17.6; DB 4; Length 169689;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 2 cattcgatcctgcgtcagtcag 25
||| ||||| ||||| |||||
Db 134076 CAATGGATCTGCTGACGACGAG 134099

RESULT 38
HSU52112 174424 bp DNA PRI 04-JAN-2001
LOCUS Homo sapiens Xq28 genomic DNA in the region of the L1CAM locus
DEFINITION containing the genes for neural cell adhesion molecule 11 (L1CAM), arginine-vasopressin receptor (AVPR2), G1 p15 (C1), ARD1 N-acetyltransferase related protein (TET2), renin-binding protein (Rbp), host cell factor 1 (HCF1), and interleukin-1 receptor-associated kinase (IRAK) genes, complete cds, and Xq281uz gene.

ACCESSION U52112
VERSION U52112.1 GI:1302657
KEYWORDS human.
SOURCE Homo sapiens


```

exon      complement(19585. .19807)
           /gene="L1CAM"
exon      complement(18
           /number=18
           complement(19990. .20060)
           /gene="L1CAM"
exon      complement(17
           /number=17
           complement(2091. .20488)
           /gene="L1CAM"
exon      complement(16
           /number=16
           complement(20735. .20845)
           /gene="L1CAM"
exon      complement(15
           /number=15
           complement(20933. .21057)
           /gene="L1CAM"
exon      complement(14
           /number=14
           complement(21237. .21393)
           /gene="L1CAM"
exon      complement(13
           /number=13
           complement(21496. .21662)
           /gene="L1CAM"
exon      complement(12
           /number=12
           complement(21776. .21887)
           /gene="L1CAM"
exon      complement(11
           /number=11
           complement(22455. .22598)
           /gene="L1CAM"
exon      complement(10
           /number=10
           complement(22738. .22869)
           /gene="L1CAM"
exon      complement(9
           /number=9
           complement(22991. .23175)
           /gene="L1CAM"
exon      complement(8
           /number=8
           complement(23323. .23434)
           /gene="L1CAM"
exon      complement(7
           /number=7
           complement(23725. .23895)
           /gene="L1CAM"
exon      complement(6
           /number=6
           complement(23992. .24114)
           /gene="L1CAM"

Query Match
Best Local Similarity 83.3%; Score 17.6; DR 97; Length 174424;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 catctgacctcgtcaggtcag 25
Db 11360 CATTTCATACCGCTCAGAGCAGC 11383

RESULT 39
AL160251 176260 bp DNA HTG 14-MAR-2001
LOCUS AL160251/C Homo sapiens chromosome 13 clone RP11-102K13, *** SEQUENCING IN
DEFINITION PROGRESS ***, 16 unordered pieces.
ACCESSION AL160251
VERSION AL160251.16 GI:13374762
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 176260)

REFERENCE
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonequest@sanger.ac.uk
On Mar 17, 2001 this sequence version replaced gi:13373899.
----- Genome Centre
Center: Sanger Centre
Center code: SC

```

```

Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: DA102K13
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 172842 bases at least Q40
Consensus quality: 173861 bases at least Q30
Consensus quality: 174344 bases at least Q20
Insert size: 174760; sum-of-contigs
Insert size: 192491; 10.8% error; agarose-fp
Quality coverage: 8.39x in Q20 bases; sum-of-contigs Quality
coverage: 8.06x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 4795: contig of 4795 bp in length
4796 4895: gap of 100 bp
4896 7081: contig of 2186 bp in length
7082 7181: gap of 100 bp
7182 14949: contig of 7768 bp in length
14950 15049: gap of 100 bp
15050 19216: contig of 4167 bp in length
19217 19316: gap of 100 bp
19317 21424: contig of 2108 bp in length
21425 21524: gap of 100 bp
21525 28685: contig of 7161 bp in length
28686 28785: gap of 100 bp
28786 88504: contig of 59719 bp in length
88505 88604: gap of 100 bp
88605 98962: contig of 10358 bp in length
98963 99062: gap of 100 bp
99063 104231: contig of 5169 bp in length
104232 104331: gap of 100 bp
104332 107452: contig of 3121 bp in length
107453 107552: gap of 100 bp
107553 123599: contig of 16047 bp in length
123600 123699: gap of 100 bp
123700 127921: contig of 4222 bp in length
127922 128021: gap of 100 bp
128022 149596: contig of 21575 bp in length
149597 149696: gap of 100 bp
149697 152985: contig of 3289 bp in length
152986 153085: gap of 100 bp
153086 158044: contig of 4959 bp in length
158045 158144: gap of 100 bp
158145 176260: contig of 18116 bp in length.
158145 176260: contig of 18116 bp in length.

FEATURES
source
1..176260
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-102K13"
/clone_id="RP11-102K13"
1..4795
/note="assembly_fragment:01005
fragment_chain:1"
4896..7081
/note="assembly_fragment:04438
fragment_chain:1"
7182..14949
/note="assembly_fragment:04451
fragment_chain:1"
15050..19216
/note="assembly_fragment:01360

```

```

misc_feature      fragment_chain:2"
                  19317..21424
                  /note="assembly-fragment:03601
fragment_chain:2"
misc_feature      21525..28685
                  /note="assembly-fragment:00700"
misc_feature      28786..88504
                  /note="assembly-fragment:00967"
misc_feature      88605..98962
                  /note="assembly-fragment:01642"
misc_feature      99063..104231
                  /note="assembly-fragment:01840"
misc_feature      104332..107452
                  /note="assembly-fragment:01987"
misc_feature      107553..123599
                  /note="assembly-fragment:02960"
misc_feature      123700..127921
                  /note="assembly-fragment:03099.0"
misc_feature      128022..149596
                  /note="assembly-fragment:03781"
misc_feature      149697..152985
                  /note="assembly-fragment:04548.0"
misc_feature      153086..158044
                  /note="assembly-fragment:05543"
misc_feature      158145..176260
                  /note="assembly-fragment:03546
                  clone_end:T7
                  vector_side:right"
BASE COUNT      45359 a 44388 c 42433 g 42578 t 1502 others
ORIGIN

```

```

Query Match      70.4%; Score 17.6; DB 79; Length 176260;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 cattgatcctgcagtcgac 25
1 ||||| ||| ||| ||| |||
Db 53420 CCTTTCATTCCTCCTCAGTCGAC 53397

```

```

RESULT 40
LOCUS      AP001446/c
DEFINITION Homo sapiens chromosome 18 clone RP11-700H19 map 18q21, WORKING
ACCESSION      AP001446.3
VERSION      AP001446.3
KEYWORDS      HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:RP11-700H19.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 186789)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
              Homo sapiens 188,789 genomic DNA of 18q21
              Published Only in Database (2000) In press
              2 (bases 1 to 188789)
              Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
              Submitted (15-MAR-2000) Masahira Hattori, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              1-15-1 Kitasato, Sagami-hara, Kanagawa 228-8555, Japan
              (E-mail:hattori@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
              Tel:81-42-778-9923, Fax:81-42-778-9924)
              On Jul 31, 2000 this sequence version replaced gi:8117320.
              ----- Genome Center
              Center: RIKEN Genomic Sciences Center (GSC)
              Center code: RIKEN
              Web site: http://hgp.gsc.riken.go.jp/
              Contact: hattori@psc.riken.go.jp
COMMENT

```

```

----- Project Information
Center project name: Humdraft18
Center clone name: RP11-700H19
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 186187 bases at least Q40
Consensus quality: 187386 bases at least Q30
Consensus quality: 187889 bases at least Q20
Insert size: 188389; sum-of-contigs
Quality coverage: 10.25x in Q20 bases; sum-of-contigs

```

```

NOTE: This is a 'working draft' sequence. It currently consists of
5 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

```

```

1 11066 contig of 11066 bp in length
11167 154535 contig of 43369 bp in length
154636 176661 contig of 22026 bp in length
176762 184717 contig of 7956 bp in length
184818 188789 contig of 3972 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 11066: contig of 11066 bp in length
* 11067 11166: gap of 100 bp
* 11167 154535: contig of 43369 bp in length
* 154536 154635: gap of 100 bp
* 154636 176661: contig of 22026 bp in length
* 176662 176761: gap of 100 bp
* 176762 184717: contig of 7956 bp in length
* 184718 184817: gap of 100 bp
* 184818 188789: contig of 3972 bp in length.

```

```

FEATURES
source
1..188789
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q21"
/clone="RP11-700H19"
1..111066
/note="assembly-fragment"
111167..154535
/note="assembly-fragment"
154636..176661
/note="assembly-fragment"
176662..184717
/note="assembly-fragment"
184818..188789
/note="assembly-fragment"
misc_feature      misc_feature
misc_feature      misc_feature
misc_feature      misc_feature
misc_feature      misc_feature
BASE COUNT      54974 a 37250 c 40336 g 55828 t 401 others
ORIGIN

```

```

Query Match      70.4%; Score 17.6; DB 82; Length 188789;
Best Local Similarity 83.3%; Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 cattgatcctgcagtcgac 25
1 ||||| ||| ||| ||| |||
Db 100853 CATTGTAACCTCCTCAGTCACG 100830

```

RESULT 41

ATCHRIV74/C
 LOCUS ATCHRIV74 194920 bp DNA PLN 16-MAR-2000
 DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 74.
 ACCESSION AL161578
 VERSION AL161578.2 GI:7269994
 KEYWORDS
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsids.
 REFERENCE 1 (bases 77867 to 116206)
 Lemnard,N., Quail,M., Harris,B., Rajandream,M.A., Barrell,B.G.,
 Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
 JOURNAL Unpublished
 REFERENCE 2 (bases 116103 to 194920)
 Brandt,P., Dose,S., Jarke,D., Scharfe,M., Schon,O., Mewes,H.W.,
 Lemcke,K. and Mayer,K.F.X.
 JOURNAL Unpublished
 REFERENCE 3 (bases 184295 to 184550)
 Volckaert,G., Gymnopoulos,B., Voelt,M., Robben,J., Mewes,H.W.,
 Lemcke,K. and Mayer,K.F.X.
 JOURNAL Unpublished
 REFERENCE 4 (bases 1 to 77970)
 Benes,V., Rechmann,S., Borkova,D., Ansoerge,W., Mewes,H.W.,
 Lemcke,K. and Mayer,K.F.X.
 JOURNAL Unpublished
 REFERENCE 5 (bases 1 to 194920)
 EU Arabidopsis sequencing project.
 JOURNAL Direct Submission
 REFERENCE Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
 Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
 lemcke@mps.biochem.mpg.de,mayer@mps.biochem.mpg.de,Project
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
 Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
 E-mail: michael.bevan@bsrc.ac.uk
 COMMENT Information on performance of analysis and a more detailed
 annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>
 this fragment has an overlap with ATCHRIV73 at the 5' end and an
 overlap with ATCHRIV75 at the 3' end.
 FEATURES
 source location/Qualifiers
 1. 194920
 /organism="Arabidopsis thaliana"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="4"
 6034..10499
 /gene="AT4g30930"
 complement(6034..6153)
 /gene="AT4g30930"
 /number=1
 complement(join(6034..6153,6271..6345,6638..6739,
 6895..6975,7060..7428,7473..7511,7837..8128,8223..8359,
 8463..9122,9444..9499,9592..9818,10309..10499))
 /gene="AT4g30930"
 complement(join(6034..6153,6271..6345,6638..6739,
 6895..6975,7060..7428,7473..7511,7837..8128,8223..8359,
 8463..9122,9444..9499,9592..9818,10309..10499))
 /gene="AT4g30930"
 /note="similarity to SP8 binding protein, Cucumis sativus,
 PIR2:Jc6203
 contains Ribosomal protein L21 signature AA724-746
 contains EST gb:AI95545.1"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB79811.1"
 /db_xref="GI:7269995"
 /translation="MGPIALKKKMTNNPSEIFPCYPAESYHMLKYVVPPLSY
 TLEBRKTNSTFLAVSSIKQNPRTKEKREKSMEDTGIDEAKTITTVKSEKVP
 EKDGISQFDEKSLGADMDLHDETVRETLGKDVQGVRENSSVEPVEDVLENET
 DSVKTVVAIVPDEVENQVETSPSLAASDSLTVTPCLSLDPAVAS7AODLPLV
 SVPTKQQRSDSPVNVRLSVTPRTPARDGYNMWKYQOKVKSFKGSRSYRCTYTE

CCAKKIECSNDSNGVVEIYVNGKLTHTHEPPRTKTSFSPREIRVTAIRPYSEDDTVVEEL
 STVPSGSDPSASTKEVCEISQTLVYRKRHCNEAVEEPEPRKROSDSVSPGKKNF
 VTHAAGDNGTICDGYRMKRYGKATKGNPHRNTYRCISACCPYRKHTEFVEMTKAY
 IITVYGVNHNDMPVPRKRHRGPPSSMLVAAPATSKRTITDOVNIPTVSSQSVGRESE
 KOSKALVDGECVCVRLMRVFDAFTSISFHLIESGTSISHGVIPKRSILRL
 PMYSHWYRSQDRCFSSNTKDDEDESEEGDEDEDESDAMEVEEYSPARK
 VEEAFETGKVMGPKLPSERLFRPEYPAIVQISHPKKSQDSITETKLCQDIN
 DKLETIKVLLLSASQTTIGRPIIPDAVYHVEEHALDEVYLIKKRRKNYRNRTRG
 HHQELTKRLRTIDTIGTEKPEKIVYKPSKEAVTEQTKALDYA"
 /number=1
 complement(6271..6345)
 /gene="AT4g30930"
 /number=2
 complement(6346..6637)
 /number=2
 complement(6638..6739)
 /gene="AT4g30930"
 /number=3
 complement(6740..6894)
 /number=3
 complement(6895..6975)
 /gene="AT4g30930"
 /number=4
 complement(6976..7059)
 /number=4
 complement(7060..7428)
 /gene="AT4g30930"
 /number=5
 complement(7429..7472)
 /number=5
 complement(7473..7511)
 /gene="AT4g30930"
 /number=6
 complement(7512..7836)
 /number=6
 complement(7837..8128)
 /gene="AT4g30930"
 /number=7
 complement(8129..8222)
 /number=7
 complement(8223..8359)
 /gene="AT4g30930"
 /number=8
 complement(8360..8462)
 /number=8
 complement(8463..9122)
 /gene="AT4g30930"
 /number=9
 complement(9123..9443)
 /number=9
 complement(9444..9499)
 /gene="AT4g30930"
 /number=10
 complement(9500..9591)
 /number=10
 complement(9592..9818)
 /gene="AT4g30930"
 /number=11
 complement(9819..10308)
 /number=11
 complement(10309..10499)
 /gene="AT4g30930"
 /number=12
 11013..12338
 /gene="AT4g30940"
 /number=1
 11013..12338
 /gene="AT4g30940"
 11013..12338
 /gene="AT4g30940"
 /note="similarity to predicted protein, Caenorhabditis
 elegans, PID:G107206

		contains EST gb:A1995535.1, N38471"		
	/codon_start=1			
	/product="putative protein"			
	/protein_id="CA679812.1"			
	/db_xref="GI:7269996"			
	/translation="MGLSDRIKIFVNGRIFFETATTLANAGRSFEGALPDENNMLIQEDLIDRNPDCFAVLIDLRLRTGDLINIPNPIPERILHKEAFYGLIDLPTAKMPDENRLHSRSVTGCIAGPGGTAIIRAGDGGCCIAHGSVIVHEPPLEEHPTINDXORNWGVMLDSGNITVALSCERLGARGDGGMGFLSSSGELRYKFQVSHNQMYSTAGALS			
	SPDSKTFTSGKGRSNEYGTGCVMDOSGKVDPDFYESPGSLDADKLQWLGSKNCLLATLPFRKDNCYISLDLFRKKNMVMSWDIGLTMAEKKVRNALAMEESNTICVNELEDGFIDLRMDGGSSVSWSSSRSLMKSMPEPCPKIALIEGOLFSSMNISVF			
	GSDVLTSLRKRSYGCSIDFSIGDRPLFALISEENVEPWETLPPIT"			
	11442..15337			
gene	/gene="A74g30950"			
	complement(join(13142..13306,13448..13562,13649..13767			
	13861..14038,14137..14259,14342..14493,14567..14710,			
	14789..15036,15309..15372,15499..15537))			
	/gene="A74g30950"			
CDS	complement(join(13142..13306,13448..13562,13649..13767			
	13861..14038,14137..14259,14342..14493,14567..14710,			
	14789..15036,15309..15372,15499..15537))			
	/gene="A74g30950"			
	/note="last exon longer than published ORF; probable			
	sequencing error genbank accession U09503			
	contains EST gb:AW004278.1, T13807, AA712148, R83991"			
	/codon_start=1			
	/product="chloroplast omega-6 fatty acid desaturase			
	(fad6)"			
	/protein_id="CAB79813.1"			
	/db_xref="GI:7269997"			
	/translation="MASRIADLSLFAFTGPQOCIPRVPKLAASAHVSFGVAVKPIDI			
	LKGTHRSRCVAVPKRRIGCIKANAVAPPSADSADEROLAESGFRIGBDLPR			
	EAVTEKJIDMTLPKEVFIEDLKALKSVLISVSYTLLGPMIAKSPPYLPLAMAMP			
	TATTFEVIGHCAKRSKSNKLVEDIYTALFIPLYVPYEMRPKRORHNAAKTMNL			
	HDTAOPVPEEEFEESPVMKRAIRIGCYPIRMWSIAMVNHNEMIKKFRSEVRVA			
	ISLACVFPMAVGPRPLIYYKGITDGWAFWMILPWLGTPHRMSTFTMVHAHLPEKI			
	ADEMNAAOQLNGTVHDYPSWTIELCHDIVNHIPHLSPRIPSYNLRAAHSIOENN			
	GKYTNLATVMMELMTIMTYCHVDKENNYIPFDRLAEESQPILEKKAMPNTYA"			
	complement(13142..13306)			
exon	/gene="A74g30950"			
	/number=1			
intron	complement(13307..13447)			
	/number=1			
exon	complement(13448..13562)			
	/gene="A74g30950"			
	/number=2			
intron	complement(13563..13648)			
	/number=2			
exon	complement(13649..13767)			
	/gene="A74g30950"			
	/number=3			
intron	complement(13768..13860)			
	/number=3			
exon	complement(13861..14038)			
Query Match	70.4%;	Score 17.6;	DB 13;	Length 194920;
Best Local	Similarity 83.3%;	Pred. No.4.5e+02;		
Matches	20;	Conservative 0;	Mismatches 4;	Indels 0;
			Gaps 0;	
Oy	1	gcattcgcgtccctcgatgata	24	
	1			
	Db 106400	GAATTGGATCTTCGTCTCATTTGCCA	106377	
RESULT	42			
LOCUS	AC023200	202733 bp	DNA	HTG
DEFINITION	Homo sapiens chromosome 8 clone RP11-48D4 map 8, WORKING DRAFT			
SEQUENCE	AC023200	45 unordered pieces.		
ACCESSION	AC023200.2	GI:7139761		
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT.			
KEYWORDS				

SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 202733)
TITLE	1 (bases 1 to 202733)
JOURNAL	Biiren,B., Linton,L., Nusbaum,C. and Lander,E.
REFERENCE	Homo sapiens chromosome 8, clone RP11-48D4
AUTHORS	2 (bases 1 to 202733)
	Biiren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Balding,J., Barra,N., Beckery,R., Beda,F., Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A., Choepl,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dekatelano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galaan,J., Gardyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Kartas,A., Klein,J., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N., Mcwan,P., McGurt,A., McKernan,K., McPheters,R., Meldrum,J., Menais,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomas,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody.M.
TITLE	Direct Submission
JOURNAL	Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Mar1, 2000 this sequence version replaced by:65957770.

*	7560	8639:	contlg of 1080 bp	in length
*	8640	8739:	gap of 100 bp	
*	8740	10075:	contlg of 1336 bp	in length
*	10076	10175:	gap of 100 bp	
*	10176	11265:	contlg of 1090 bp	in length
*	11266	11365:	gap of 100 bp	
*	11366	12553:	contlg of 1188 bp	in length
*	12554	12653:	gap of 100 bp	
*	12654	13991:	contlg of 1338 bp	in length
*	13992	14091:	gap of 100 bp	
*	14092	15357:	contlg of 1266 bp	in length
*	15358	15457:	gap of 100 bp	
*	15458	16488:	contlg of 1031 bp	in length
*	16489	16588:	gap of 100 bp	
*	16589	18103:	contlg of 1515 bp	in length
*	18104	18203:	gap of 100 bp	
*	18204	19604:	contlg of 1401 bp	in length
*	19605	19704:	gap of 100 bp	
*	19705	21252:	contlg of 1548 bp	in length
*	21253	21352:	gap of 100 bp	
*	21353	22624:	contlg of 1272 bp	in length
*	22625	22724:	gap of 100 bp	
*	22725	24739:	contlg of 2015 bp	in length
*	24740	24839:	gap of 100 bp	
*	24840	26507:	contlg of 1668 bp	in length
*	26508	26607:	gap of 100 bp	
*	26608	28260:	contlg of 1653 bp	in length
*	28261	28360:	gap of 100 bp	
*	28361	28482:	contlg of 1122 bp	in length
*	29483	29582:	gap of 100 bp	
*	29583	31417:	contlg of 1835 bp	in length
*	31418	31517:	gap of 100 bp	
*	31518	33784:	contlg of 2267 bp	in length
*	33785	33884:	gap of 100 bp	
*	33885	358218:	contlg of 2334 bp	in length
*	36219	36318:	gap of 100 bp	
*	36319	37338:	contlg of 1020 bp	in length
*	37339	37438:	gap of 100 bp	
*	37439	39380:	contlg of 1942 bp	in length
*	39381	39480:	gap of 100 bp	
*	39481	40995:	contlg of 1515 bp	in length
*	40996	41095:	gap of 100 bp	
*	41096	44026:	contlg of 2931 bp	in length
*	44027	44126:	gap of 100 bp	
*	44127	46095:	contlg of 1969 bp	in length
*	46096	46195:	gap of 100 bp	
*	46196	48391:	contlg of 2196 bp	in length
*	48392	48491:	gap of 100 bp	
*	48492	51248:	contlg of 2757 bp	in length
*	51249	51348:	gap of 100 bp	
*	51349	54496:	contlg of 3148 bp	in length
*	54497	54596:	gap of 100 bp	
*	54597	59348:	contlg of 4752 bp	in length
*	59349	59448:	gap of 100 bp	
*	59449	62763:	contlg of 3315 bp	in length
*	62764	62863:	gap of 100 bp	
*	62864	67726:	contlg of 4863 bp	in length
*	67727	67826:	gap of 100 bp	
*	67827	74038:	contlg of 6212 bp	in length
*	74039	74138:	gap of 100 bp	
*	74139	79699:	contlg of 5557 bp	in length
*	79696	79795:	gap of 100 bp	
*	79796	87048:	contlg of 7253 bp	in length
*	87149	87148:	gap of 100 bp	
*	87149	96461:	contlg of 9313 bp	in length
*	96462	96561:	gap of 100 bp	
*	96562	107284:	contlg of 10723 bp	in length
*	107285	107284:	gap of 100 bp	
*	107285	118979:	contlg of 12495 bp	in length
*	119880	119979:	gap of 100 bp	
*	119980	135672:	contlg of 15693 bp	in length
*	135673	135772:	gap of 100 bp	
*	135773	154230:	contlg of 18458 bp	in length

FEATURES	* 154231 154330: gap of 100 bp
	* 154331 172603: contig of 18273 bp in length
	* 172604 172703: gap of 100 bp
	* 172704 202733: contig of 30030 bp in length.
FEATURES	Location/Qualifiers
source	1..202733
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="8"
	/map="8"
	/clone="RP11-48D4"
	/clone_11b="RPC1-11 Human Male BAC"
	1..1103
misc_feature	/note="assembly_fragment"
	1204..2241
misc_feature	/note="assembly_fragment"
	2342..3589
misc_feature	/note="assembly_fragment"
	3690..4933
misc_feature	/note="assembly_fragment"
	5034..6204
misc_feature	/note="assembly_fragment"
	6305..7459
misc_feature	/note="assembly_fragment"
	7560..8639
misc_feature	/note="assembly_fragment"
	8740..10075
misc_feature	/note="assembly_fragment"
	10176..11265
misc_feature	/note="assembly_fragment"
	11366..12553
misc_feature	/note="assembly_fragment"
	12654..13991
misc_feature	/note="assembly_fragment"
	14092..15357
misc_feature	/note="assembly_fragment"
	15458..16488
misc_feature	/note="assembly_fragment"
	16589..18103
misc_feature	/note="assembly_fragment"
	18204..19604
misc_feature	/note="assembly_fragment"
	19705..21252
misc_feature	/note="assembly_fragment"
	21353..22624
misc_feature	/note="assembly_fragment"
	22725..24739
misc_feature	/note="assembly_fragment"
Query Match	70.4%: Score 17.6; DB 67; Length 202733;
Best Local Similarity	83.3%: Pred. No.4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;	
QY	2 catlccgactcctgcacgaatgacg 25
Db 180816	CATGTCGATCCTATTCAGTTCAG 180793
RESULT 43	
CNS01R12/c	CNS01R12 208103 bp DNA HTG 19-FEB-2001
LOCUS	Homo sapiens chromosome 14 clone R-1077A2, *** SEQUENCING IN
DEFINITION	PROGRESS ***, in ordered pieces.
ACCESSION	AL162831.4 GI:13016592
VERSION	AL162831.4
KEYWORDS	HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 208103)
AUTHORS	Genoscope.
TITLE	Direct Submission

JOURNAL

COMMENT

Submitted (19-FEB-2001) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
 Web : www.genoscope.cns.fr)
 On Feb 20, 2001 this sequence version replaced gi:8919853.
 ----- Genome Center
 Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: Sequef@genoscope.cns.fr

IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in progress
 and the release of this data is based on the understanding that the
 sequence may change as work continue. The sequence may be
 contaminated with foreign sequence from E.coli, yeast, vector,
 phage, etc. even if efforts are made to eliminate these
 contaminating sequences. The following BAC sequence is oriented
 from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : R-108M12 (AC-AL137100)
 Downstream BAC (overlapping the SP6 end) : R-1085N6 -----
 Summary Statistics
 Assembly program: Phrap: version 2.0
 Quality coverage: 7.88x in Q20 bases; sum-of-contigs

 Overall quality chart :
 Range : bases
 0 :
 1 - 9 : 31
 10 - 19 : 201
 20 - 29 : 559
 30 - 39 : 1888
 40 - 49 : 8300
 50 - 59 : 19871
 60 - 69 : 12111
 70 - 79 : 17261
 80 - 89 : 41493
 90 - 99 : 106388

 Percentage of bases with a quality value >= 40 : 98 %
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 Location/Qualifiers
 1. 208103
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone="R-1077A2"
 /clone_1lb="R-1077A2"
 72007..72164
 /note="matching EMBL:AA065110
 Rhab:RH47643
 dbSTS:STS40702
 Identified using the e-PCR software (G. Schuler)"
 94217..94316
 /note="matching EMBL:G21898
 Rhab:RH53515
 dbSTS:STS5867
 Identified using the e-PCR software (G. Schuler)"
 140405..140500
 /note="matching EMBL:G42446
 Rhab:RH101647
 dbSTS:STS69204
 Identified using the e-PCR software (G. Schuler)"
 145757..145906
 /note="matching EMBL:HB5844
 Rhab:RH53626
 dbSTS:STS19601
 Identified using the e-PCR software (G. Schuler)"

STS

STS

161830..162062
 /note="matching EMBL:R08155
 Rhab:RH68366
 dbSTS:STS48273
 Identified using the e-PCR software (G. Schuler)"
 161902..162023
 /note="matching EMBL:R17804
 Rhab:RH53746
 dbSTS:STS4425
 Identified using the e-PCR software (G. Schuler)"
 BASE COUNT 62254 a 42380 c 42335 g 61134 t
 ORIGIN

Query Match 70.4% Score 17.6; DB 84; Length 208103;
 Best Local Similarity 83.3% Pred. No. 4.5e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 cattgatcctcgtcaggtcag 25
 ||||| | ||||| || |||||
 Db 156225 CATTGGGTCTCGCAGCTGCAG 156202

RESUL.F 44
 AE003814

LOCUS 260713 bp DNA INV 04-OCT-2000
 DEFINITION Drosophila melanogaster genomic scaffold 142000013386047 section 29
 of 52, complete sequence.
 ACCESSION AE003814 AE002787
 VERSION AE003814.2 GI:10727549
 KEYWORDS
 SOURCE
 ORGANISM
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
 AUTHORS

1 (bases 1 to 260713)
 Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
 Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
 George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
 Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
 Brandon,R.C., Rogers,Y.H., Blazer,V., Buzze,R.G., Champe,M., Pfeiffer,B.D.,
 Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gaber
 Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
 Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
 Bayendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
 Benos,P.V., Bernier,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
 Borchan,M.R., Bouck,J.J., Brokstein,P., Brotlier,P., Burris,K.C.,
 Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
 Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
 Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
 Dodson,P., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
 Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriere,S.,
 Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S.,
 Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
 Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
 Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
 Wel,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpn,G.H., Ke,Z.,
 Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
 Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Maltel,B., McIntosh,T.C.,
 McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
 Morris,J.J., Moshrefi,A., Mount,S.N., Moy,M., Murphy,B., Murphy,L.,
 Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
 Nusskern,D.R., Paclebl,J.M., Palazolo,M., Pittman,G.S., Pan,S.,
 Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
 Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
 Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
 Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
 Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,

Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of *Drosophila melanogaster*
JOURNAL
MEDLINE
2 (bases 1 to 260713)
20196006
AUTHORS
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
JOURNAL
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
On Oct 9, 2000 this sequence version replaced gi:7303162.
FEATURES
source
1. 260713
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
join(<878..988,1057..1277,1507..1663,2021..2134,
2271..2491,2709..2800,3305..3351,3416..3761,3994..4088,
5335..5850)
/gene="CG13940"
/product="CT33486"
/db_xref="FLYBASE:FBan0013940"
/db_xref="FLYBASE:FBgn0033930"
/evidence=not_experimental
<878..5850
/gene="CG13940"
/map="50F6-50F7"
/db_xref="FLYBASE:FBan0013940"
/db_xref="FLYBASE:FBgn0033930"
/evidence=not_experimental
join(878..988,1057..1277,1507..1663,2021..2134,2271..2491,
2709..2800,3305..3351,3416..3761,3994..4088,5335..5850)
/gene="CG13940"
/note="CG13940 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0013940"
/db_xref="FLYBASE:FBgn0033930"
/evidence=not_experimental
/protein_id="AAF58252.1"
/db_xref="GI:7303186"
/translation="MAISVPIGLPHRNFLSYNERYVYQPEHYVKKYPIILMGDFED
SYLTPTTGREAQRHQCQTDWKIEGNINSTSNNSSTKASREKRGJLLMSRVFY
AMLRDLRSPSSKDEHLNENYQAEQMDREDEQESTYTKSCDNLIDLSVLSRLSL
SDIIFAIVPLELPHRNVSYPFANPLANMEKWTIPQNGPISEEVVDEIDTE
TDRKLAGCQNCYKKEENAGSEFEVETEVLPQEKVRSLLFRSNTYITPDKLSP
SSSEEDLPLRYQAEHDGNGCHVEPVSTDLPRSMOKPPELRHNGTDESLLS
DHGHNDNDVYDDYDDQNGNKHKHHHPHNDKRRKSKRPQSVIKRPOKRRP
KHKHKKRHSKPRPEVEDDYKDFQGFQDGMQGESVGRDQIIPALGMKNYLSVT
TSGNMVVGMAATAEICNCRYECLRYHMDLDDNKKRPDELIMTSLSPNGDK
YGAARFAHETGALGSRHADPELLYSNOYADKMGMASSCTLPYAMLHACGMVYLTA
NCPRENIIDDPKCNLSQKLISCTKKIDETNALKKRDELDLIDNGCHIDISGNSILM
ACFLTLMIAKFISDH"
complement(join(<8403..8617,8675..8831,8891..>8974))
/gene="CG13939"
/product="CT33485"
/db_xref="FLYBASE:FBan0013939"
/db_xref="FLYBASE:FBgn0033931"
/evidence=not_experimental
complement(<8403..>8974)
/gene="CG13939"
/map="50F6-50F8"
/db_xref="FLYBASE:FBan0013939"
/db_xref="FLYBASE:FBgn0033931"
/evidence=not_experimental
complement(join(8403..8617,8675..8831,8891..8974))
/gene="CG13939"
/note="CG13939 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0013939"
/db_xref="FLYBASE:FBgn0033931"
/evidence=not_experimental
/protein_id="AAF58251.1"
/db_xref="GI:7303187"

/translation="MGDVPQCKHYVSGLSKANSKYPSYAHLCPCDIYRENGAMVNG
KIKYRKYVEEHRHRODEIYSHIYQSESCISYNGKMGKLNIESKYVLPHGCS
PAGITVACVNAETFLNCPQOMKMKERKNLAKQFAEQCNLPHVLPSS"
join(11956..12486,16004..16301,16526..16724,16841..16934,
17080..17254,17313..17386,17454..17564,17793..17981,
18049..18164,18413..18469,18533..18740,18806..>18832)
/gene="CG8422"
/product="CT24513"
/db_xref="FLYBASE:FBan0008422"
/db_xref="FLYBASE:FBgn0033932"
<11956..>18832
/gene="CG8422"
/map="50F6-50F9"
/db_xref="FLYBASE:FBan0008422"
/db_xref="FLYBASE:FBgn0033932"
join(16037..16301,16526..16724,16841..16934,17080..17254,
17313..17386,17454..17564,17793..17981,18049..18164,
18413..18469,18533..18740,18806..18832)
/gene="CG8422"
/note="CG8422 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0008422"
/db_xref="FLYBASE:FBgn0033932"
/protein_id="AAF58250.1"
/db_xref="GI:7303186"
/translation="MSDHHIDSYNAGSDPLDLNHLNDIGESVEIQLCYEHIEAS
TYGNDSCGLTQPSDILCWPTARGTTLAIOCMELQIHYDSSKNRFRCHANGTWE
KTYNDACAHLPAPESVEFEVYIEPLTIITYITLTSLYSIALILVAFYKELRCL
KNTIHANLFPTVIMSALFWILLISVOISIRKSGVSCJALITLHFTFLNFRMLVYG
LNYLMVVRPFGSGLNRENIVASIGMGSPALVVYVAVANSILVYTSPEKYEYINCG
MQEYHVIDIYQPYCAVILINLFTLRIMVWLITRKLSANVETROYRKAKALVLI
PLFGITVYLVLAGSESGLMGMFAVLAVALITSGQSVSLFYCLNSEVYRNALRHII
STWRTRTIOLNOMNRYTTKSPSGGSGPRAESMRPLTSYRGKRSQVSATTTTL
VQGNRPLSHRGSNNALHTMTPLAANMSSGSLTSVMPRAISPLMRGLENSY"
<24631..>25845
/gene="CG10104"
/product="CT28441"
/db_xref="FLYBASE:FBan0010104"
/db_xref="FLYBASE:FBgn0033933"
<24631..>25845
/gene="CG10104"
/map="50F6-51A1"
/db_xref="FLYBASE:FBan0010104"
/db_xref="FLYBASE:FBgn0033933"
24631..25845
/gene="CG10104"
/note="CG10104 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0010104"
/db_xref="FLYBASE:FBgn0033933"
/protein_id="AAF58249.1"
/db_xref="GI:7303185"
/translation="WMLVLSLPLVFLIPVOPQHPVSCQLDLYRPLRRPSAHRRE
KLGIIMDRILRYAEVSHFRGEMNSAVKSPLSLVYDAQFGPTTIGTPQTRVIF
DTGSNLLWPSATCSTWVACRVHNRYPAKSTSHOYGDHFAIHYGSGSLGSLSD
TVRVAGLELRDQTEATEMPQPIPLAKFGIGLAVRYSIMQRIKPPFANMOGL
LTKPIPSYVLSRNGEKDGAIFEGSNHNTYNTQVYSHRAYVOYKMDASVIRNL
ETCQGGCEVYITDGTSTFLALPDDAIIINESIGTPSPSGFQVLPQCSVPLPTTF
LGGRRFLESHEHYFRDIYQDRITCSSAFINAVDLPSPGPMILLDVFLGXYTFPM
EHRHIFADARS"
complement(join(<44052..44259,44566..45125,45184..>45252))
/gene="CG17385"
/product="CT33480"
/db_xref="FLYBASE:FBan0017385"
/db_xref="FLYBASE:FBgn0033934"
complement(<44052..>45252)
/gene="CG17385"
/map="50F9-51A1"
/db_xref="FLYBASE:FBan0017385"
/db_xref="FLYBASE:FBgn0033934"
complement(join(44052..44259,44566..45125,45184..45252))
/gene="CG17385"
/note="CG17385 gene product"

```

/codon_start=1
/db_xref="FLYBASE:FBan0017385"
/db_xref="FLYBASE:FBan0033934"
/protein_id="AAFS8248.1"
/db_xref="GI:7303184"
/translation="MAGMEVFDEFVANQFSCRCRDRTFNSKRQTLHROEVNHNKT

Query Match      70.4%: Score 17.6; DB 5; Length 260713;
Best Local Similarity 83.3%: Pred. No. 4.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 2 cattcgatccctcagtgag 25
      ||||| ||||| |||||
Db 240134 CAATTGATCTCTGTCAGAGCAG 240157

RESULT 45
AL499627 LOCUS
DEFINITION Human DNA sequence from clone RP13-379024 on chromosome 20 contains
a gene for a novel protein similar to transcription factor GATA-5,
ESTs, STSs, GSSs and two Cpg islands, complete sequence.
AL499627 AL499627.23 GI:13559068
HTG: CPG island; GATA-5; transcription factor.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 59999)
AUTHORS Matthews, L.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK E-mail enquiries: humgen@esanger.ac.uk
requests: clone@esanger.ac.uk
On Apr 5, 2001 this sequence version replaced gi:13446506.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats: all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; SW:
SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP13-379024 is from the library RPCT-13.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
RP13-379024. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP13-379024 is at 59999 in this
sequence. The true right end of clone RP5-908M14 is at 100 in this
sequence.

FEATURES
Source
1. 59999
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"

/codon="RP13-379024"
/clone.lib="RPCT-13.2"
/complement(1..95)
/note="match: STS: Em: HS908M14T"
1020..1119
/note="2 copies 50 mer 91% conserved"
1026..1121
/note="4 copies 24 mer 79% conserved"
3539..3867
/note="Aluio repeat: matches 3..306 of consensus"
3939..4116
/note="Aluio repeat: matches 54..220 of consensus"
5954..6007
/note="2 copies 27 mer 96% conserved"
6334..6456
/note="MER5A repeat: matches 1..125 of consensus"
6931..7062
/note="4 copies 33 mer 84% conserved"
/complement(7115)
/clone="DB379024.1"
/complement(join(7115..8609,8958..9082,9452..9539,
10044..10169,17021..17196,18617..19251))
/note="match: cDNAs: Em: U84725
match: ESTs: Em: H80372 Em: H80385 Em: W00677 Em: A1632657
Em: N72525 Em: H80289 Em: AA127875 Em: AA127913 Em: AW770386
Em: AA972923 Em: A1685293 Em: AA534131 Em: AW973240
Em: AA533669 Em: AA127923 Em: N53566 Em: AW511251 Em: A1174441"
/evidence="not experimental"
/product="DB379024.1 (novel) protein similar to
transcription factor GATA-5"
/clone="DB379024.1"
/clone="DB379024.1"
/clone="DB379024.1"
7724..8137
/note="match: STS: Em: G25060"
7724..8137
/note="match: STS: Em: G29730"
7764..7900
/note="MER91B repeat: matches 1..124 of consensus"
/complement(join(8454..8609,8958..9082,9452..9539,
10044..10169,17021..17196,18617..19139))
/clone="DB379024.1"
/note="match: proteins: Tr: O16365 Sw: P43692 Wp: P52C12
Tr: O97YX3 Sw: P19212 Sw: P23772 Sw: P23825 Sw: O91428
Tr: O97YB8 Sw: O13415 Tr: O77137 Sw: P28515 Sw: P43695
Sw: O91678 Sw: O92908 Sw: O91677 Sw: P70005 Sw: P23773
Sw: P23770 Sw: P17679 Tr: O90410 Sw: P17678 Sw: P43429"
/codon_start=1
/evidence="not experimental"
/product="DB379024.1 (novel) protein similar to
transcription factor GATA-5"
/transcript_id="CAC36001.1"
/db_xref="GI:13559069"
/translation="MYOSIALASBPQAAVADSGSFLHAPGAGSPMFPVAPVPSMIS
LYSGSPPEPELAAPGMAOTATADSSAGSPSPHPPAHPPGATAPFAPHSPPG
GSGSAGGDSAGVAGALLPPEQAPAPAGRGVGSATYVAPVAPVAPVAPVAPV
GSLVHGLGGRRTPTVSDLEFPPEGRPCVNCAGLSTPLMRDGTGYLCAACLYTK
MGVAVRPLVRFQKRLLSSRRAGLCCCTNCHTNTYTLWRNSGEPEVCNACGLYMKLHY
PPLAMKRESIQTRKRKPTTAKAGSGSTRNASASPSAASVADSSNAASKAPSLA
SPVCPGSPMAPQASQEDDSIAPGLHLEKPEPEDFAPPTAPSPQAGLRGALROEAWC
ALALA"
9836..10334
/note="Cpg island"
/evidence="not experimental"
10311..10434
/note="MIR repeat: matches 62..188 of consensus"
11461..11560
/note="2 copies 50 mer 86% conserved"
11596..11651
/note="2 copies 28 mer 94% conserved"
11672..11727

```

repeat_region	/note="2 copies 28 mer 94% conserved"
	11744. .11799
repeat_region	/note="2 copies 28 mer 94% conserved"
	11820. .1875
repeat_region	/note="2 copies 28 mer 96% conserved"
	11892. .1947
repeat_region	/note="2 copies 28 mer 96% conserved"
	12774. .12958
repeat_region	/note="MERS3 repeat: matches 10. .189 of consensus"
	14602. .14773
repeat_region	/note="MIR repeat: matches 37. .214 of consensus"
	15051. .15342
repeat_region	/note="AluB repeat: matches 7. .299 of consensus"
	15351. .15663
repeat_region	/note="AluB repeat: matches 1. .296 of consensus"
	17210. .17291
misc_feature	/note="2 copies 41 mer 90% conserved"
	17861. .20459
	/note="Cpg island"
	/evidence=not_experimental
misc_feature	19352. .20175
	/note="Single clone region. Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
misc_feature	19448. .19645
misc_feature	/note="match: GSS: Em:A0938920"
repeat_region	19641. .19712
repeat_region	/note="3 copies 24 mer 84% conserved"
	19642. .19713
misc_feature	/note="36 copies 2 mer cc 70% conserved"
	complement(20394. .20581)
repeat_region	/note="match: STS: Em:L30262"
	20489. .20936
repeat_region	/note="117 copies 4 mer ttcc 66% conserved"
	20492. .20722
repeat_region	/note="7 copies 33 mer 67% conserved"
	20496. .20959
repeat_region	/note="29 copies 16 mer 65% conserved"
	20500. .20950
repeat_region	/note="11 copies 41 mer 62% conserved"
	20504. .20951
repeat_region	/note="16 copies 28 mer 66% conserved"
	20724. .20789
repeat_region	/note="2 copies 33 mer 93% conserved"
	20827. .20925
repeat_region	/note="3 copies 33 mer 76% conserved"
	20831. .20920
misc_feature	/note="15 copies 6 mer ttcttc 67% conserved"
	20933
repeat_region	/note="Random repeat. Forced join. Gap size estimated to be approximately 110bp by restriction digest data."
	20958. .21234
repeat_region	/note="AluX repeat: matches 9. .294 of consensus"
	22223. .22386
repeat_region	/note="MIR repeat: matches 52. .226 of consensus"
	22934. .23051
repeat_region	/note="L2 repeat: matches 2349. .2473 of consensus"
	23253. .23314
repeat_region	/note="MIR repeat: matches 122. .183 of consensus"
	23676. .23871
repeat_region	/note="7 copies 28 mer 62% conserved"
	23687. .23851
repeat_region	/note="5 copies 33 mer 84% conserved"
	23689. .23877
repeat_region	/note="7 copies 27 mer 64% conserved"
	23697. .23846
repeat_region	/note="3 copies 50 mer 74% conserved"
	23741. .23868
repeat_region	/note="8 copies 16 mer 67% conserved"
	25184. .25344
repeat_region	/note="MIR repeat: matches 21. .190 of consensus"
	27265. .27342

Query Match	69.6%;	Score 17.4;	DB 91;	Length 59999;
Best Local Similarity	94.7%;	Pred. No. 5.4e+02;		
Matches 18; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;
QY 5 ttgatcctctgaaggtgc 23				
Db 12630 ttctatcctctgacgggtcc 12648				

```
Search completed: October 9, 2001, 12:14:30
Job time: 3985 sec
```

Wed Oct 10 07:44:51 2001

us-09-396-196f-8.std.rge

Page 38

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 11:39:28 ; Search time 470.56 Seconds
(without alignments)
33.359 Million cell updates/sec

Title: US-09-396-196f-8
Perfect score: 25
Sequence: 1 gcatttcgacccgctcagtgacg 25

Scoring table:
IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 730101 seqs, 313950809 residues
Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDSI/gcgcdata/geneseq/NA1980.DAT:*
2: /SIDSI/gcgcdata/geneseq/NA1981.DAT:*
3: /SIDSI/gcgcdata/geneseq/NA1982.DAT:*
4: /SIDSI/gcgcdata/geneseq/NA1983.DAT:*
5: /SIDSI/gcgcdata/geneseq/NA1984.DAT:*
6: /SIDSI/gcgcdata/geneseq/NA1985.DAT:*
7: /SIDSI/gcgcdata/geneseq/NA1986.DAT:*
8: /SIDSI/gcgcdata/geneseq/NA1987.DAT:*
9: /SIDSI/gcgcdata/geneseq/NA1988.DAT:*
10: /SIDSI/gcgcdata/geneseq/NA1989.DAT:*
11: /SIDSI/gcgcdata/geneseq/NA1990.DAT:*
12: /SIDSI/gcgcdata/geneseq/NA1991.DAT:*
13: /SIDSI/gcgcdata/geneseq/NA1992.DAT:*
14: /SIDSI/gcgcdata/geneseq/NA1993.DAT:*
15: /SIDSI/gcgcdata/geneseq/NA1994.DAT:*
16: /SIDSI/gcgcdata/geneseq/NA1995.DAT:*
17: /SIDSI/gcgcdata/geneseq/NA1996.DAT:*
18: /SIDSI/gcgcdata/geneseq/NA1997.DAT:*
19: /SIDSI/gcgcdata/geneseq/NA1998.DAT:*
20: /SIDSI/gcgcdata/geneseq/NA1999.DAT:*
21: /SIDSI/gcgcdata/geneseq/NA2000.DAT:*
22: /SIDSI/gcgcdata/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	20	AA01303
2	25	100.0	1084	10	AA01329
3	25	100.0	1121	7	AA060496
4	25	100.0	5872	15	AA062386
5	17.6	70.4	565	21	AA074619
6	17.2	68.8	241	21	AA036290
7	17.2	68.8	393	20	AA080637
8	17	68.0	1542	22	AA082667
9	17	68.0	1782	22	AA083079
10	17	68.0	2254	19	AA065382
11	17	68.0	3234	12	AA015019

C 12	16.6	66.4	34094	20	AA030163
C 13	16.2	64.8	385	21	AA056248
C 14	16.2	64.8	393	21	AA057313
C 15	16.2	64.8	865	18	AA030823
C 16	16.2	64.8	882	19	AA098665
C 17	16.2	64.8	1702	19	AA029636
C 18	16.2	64.8	1702	19	AA037384
C 19	16.2	64.8	14736	19	AA052304
C 20	16	64.0	1045	18	AA059977
C 21	16	64.0	1045	18	AA052016
C 22	16	64.0	1137	20	AA036567
C 23	16	64.0	1137	21	AA022300
C 24	16	64.0	1152	21	AA065913
C 25	16	64.0	1152	21	AA039397
C 26	16	64.0	1792	20	AA059380
C 27	16	64.0	2567	20	AA052889
C 28	16	64.0	3708	13	AA023963
C 29	16	64.0	3904	21	AA053179
C 30	16	64.0	3954	18	AA059975
C 31	16	64.0	3954	20	AA020015
C 32	16	64.0	4928	16	AA094735
C 33	16	64.0	4928	17	AA040205
C 34	16	64.0	11854	20	AA013243
C 35	16	64.0	20757	20	AA020599
C 36	16	64.0	143068	21	AA021105
C 37	16	64.0	143068	21	AA021272
C 38	16	64.0	143068	21	AA034983
C 39	16	64.0	143068	21	AA035150
C 40	16	64.0	149412	21	AA035151
C 41	16	64.0	152740	21	AA021273
C 42	16	64.0	534720	19	AA030458
C 43	15	63.2	536165	19	AA030459
C 44	15	63.2	220	22	AA028005
C 45	15.8	63.2	220	22	AA028044

ALIGNMENTS

RESULT 1
AA01303 standard; DNA; 1041 BP.
ID AA01303
XX
AC AA01303;
XX
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (BioB) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthetase; biotin production; vitamin H; BioB; ss.
XX
OS Escherichia coli.
XX
PN US5869719-A.
XX
PD 09-FEB-1999.
XX
PF 30-APR-1997; 97US-0846338.
XX
PR 30-APR-1997; 97US-0846338.
XX
PR 08-MAR-1995; 95US-0401068.
XX
PA (NOVS) NOVARTIS FINANCE CORP.
XX
PI Patton DA;
XX
DR WPI; 1999-152902/13.
XX
PT P-PSDB; AA073906.
XX
PT Transgenic plants with high biotin levels - transformed with DNA
PT encoding diaminopelargonic acid amino-transferase or biotin
PT synthase

Complete nucleoid
Pinus radiata tran
Pinus radiata tran
Streptococcus pneu
DNA encoding a S.
S. pneumoniae deri
Streptococcus pneu
Streptococcus pneu
5' untranslated re
D. melanogaster tl
Human EDG-3 coding
Human EDG3b polyP
E. coli proliferat
Escherichia coli m
Aspergillus nidula
Human prostate tum
Probe Bly2. Bos t
Streptococcus bov1
Full length type P
D. melanogaster tl
Lycopene cyclase g
Cyanothece sp. 116
Enterococcus faeca
Polynucleotide seq
Human low adenosin
Human low adenosin
Human low adenosin
Human low adenosin
Human low adenosin
Human low adenosin
Rhizobium species
Rhizobium species
Human TR13 coding
Human TR13 coding

XX Example 2; Column 37-40; 34pp; English.

CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC be used in the transgenic plant of the invention. The transgenic plant,
CC plant cell or plant tissue is transformed with a chimeric gene encoding
CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC produces more biotin than a non-transgenic plant, cell or tissue. The
CC plant is used as an improved dietary source of biotin (vitamin H) for
CC humans or animals.

XX Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0074;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatctgacatcctgcagtgacg 25
|||||
DB 99 gcattcgatcctcctgcagtgacg 123

RESULT 2

AA091329 standard; DNA; 1084 BP.

AC AA091329;

DT 15-FEB-1990 (first entry)

DE E.coli Bio B gene.

DE E.coli Bio B gene; biotin.

OS Escherichia coli.

Key Location/Qualifiers

FT CDS 24..1064

FT /*tag=a

PN GB2216530-A.

PD 11-OCT-1989.

PF 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

PR 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISH.

PI Pearson BM, McKee RA;

DR WPI; 1989-295085/41. P-PSDB P91392

XX Plasmid contg. gene(s) for expression of biotin synthetase enzymes
XX - derived from E.coli and capable of replication and expression in other
XX microorganisms, esp. yeast.

PS Table 3; page 33-4; 52pp; English.

CC The gene can be used in a plasmid for expression of enzymes of the biotin
CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
CC Lactobacillus. Insertion of bio B improves biotin yields in
CC microorganisms which export biotin, or enables growth in media contg.
CC little or no biotin of organisms unable to synthesise biotin for their
CC own use.

XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.0075;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcatctgacatcctgcagtgacg 25
|||||
DB 122 gcattcgatcctcctgcagtgacg 146

RESULT 3

AA060496 standard; DNA; 1121 BP.

AC AA060496;

DT 17-OCT-1991 (first entry)

DE Sequence encoding biotin synthesising enzyme.

KW Biotin synthetic enzyme; E.coli; desthiobiotin; ds.

Key Location/Qualifiers

FT CDS 42..1082

PN JP61149091-A.

PD 07-JUL-1986.

PF 24-DEC-1984; 84JP-0272605.

PR 24-DEC-1984; 84JP-0272605.

PA (NIPS) NIPPON SODA KK.

DR WPI; 1986-216622/33.

DR P-PSDB; AAP60536.

PT Double stranded DNA encoding biotin synthesising enzyme -
PT comprises transformed mutant E.coli strain contg. cyclic doubled
PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

PS Disclosure; Page 534; 23pp; Japanese.

CC The sequence may be expressed by a transformed E.coli host, cultured
CC in a medium containing desthiobiotin.

CC Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

KM dethiobiotin synthase; DAPA synthase;
 KM S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
 KW seborrhoea; dermatitis; ds.
 OS Escherichia coli DSM498.
 XX
 FH Key Location/Qualifiers
 FT promoter 1..96
 FT /tag= a
 FT /function= "promoter plac"
 FT /evidence= EXPERIMENTAL
 FT 23..28
 FT -35_signal
 FT /tag= b
 FT /standard_name= "promoter plac"
 FT 45..50
 FT /tag= c
 FT /evidence= EXPERIMENTAL
 FT /standard_name= "promoter plac"
 FT 105..109
 FT /tag= d
 FT /evidence= EXPERIMENTAL
 FT /standard_name= "bioB RBS no. 9"
 FT 117..1157
 FT CDS
 FT /tag= e
 FT /product= "biotin synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioB"
 FT /number= 1
 FT 1141..1146
 FT /tag= f
 FT /standard_name= "bioF RBS"
 FT 1154..2311
 FT /tag= g
 FT /EC_number= 2.3.1.47
 FT /product= "KAPA synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioF"
 FT /number= 2
 FT /standard_name= "8-amino-7-oxononanoate synthase"
 FT 2284..2288
 FT /tag= h
 FT /standard_name= "biOC RBS"
 FT 2295..3050
 FT CDS
 FT /tag= i
 FT /function= "involved in pimeoloyl-CoA synthesis"
 FT /product= "protein"
 FT /gene= "biOC"
 FT /number= 3
 FT 3030..3033
 FT /tag= j
 FT /standard_name= "biOD RBS"
 FT 3043..3753
 FT /tag= k
 FT /EC_number= 6.3.3.3
 FT /product= "DTB synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "biOD15"
 FT /number= 4
 FT /standard_name= "dethiobiotin synthase"
 FT 3712..3750
 FT /tag= l
 FT /note= "biOD15 substitution"
 FT 3742..3746
 FT RBS
 FT /tag= m
 FT /standard_name= "bioA RBS"
 FT 3750..5039
 FT /tag= n
 FT /EC_number= 2.6.1.62
 FT /product= "DAPA synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioA"
 FT /number= 5
 FT /standard_name= "S-adenosyl-L-methionine: 8-amino-

FT RBS 5088..5093
 FT /tag= o
 FT /standard_name= "ORF1 RBS"
 FT 5098..5574
 FT CDS
 FT /tag= p
 FT /function= "unknown, involved in biotin synthesis"
 FT /product= "protein"
 FT /evidence= EXPERIMENTAL
 FT /gene= "ORF1"
 FT /number= 6
 FT 5583..5644
 FT terminator
 FT /tag= q
 FT /standard_name= "rho-independent transcriptional terminator"
 FT 5583..5605
 FT /tag= r
 FT stem_loop
 FT 5583..5605
 FT /tag= r
 FT WO9408023-A.
 FT 14-APR-1994.
 FT 01-OCT-1993; 93WO-EP02688.
 FT 02-OCT-1992; 92CH-0003124.
 FT 15-JUL-1993; 93CH-0002134.
 FT (LONZ) LONZA AG.
 FT Birch O, Braas J, Fuhrmann M, Shaw N;
 FT WPI: 1994-135587/16.
 FT P-PDB: AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
 FT Biotechnological biotin prodn. using enterobacterial biotin-gene
 FT - providing vitamin H in high yield
 FT
 FT Claim 1: Fig 6, page 47-55 and 60-65; 92pp; German.
 FT
 FT The sequence is derived from plasmid pB030A-15/9 contg. the
 FT bioB, bioF, biOC, biOD and bioA genes responsible for biosynthesis
 FT of biotin, arranged in a transcription unit. Microorganisms
 FT contg. these DNA fragments or plasmids may be used in the prodn.
 FT of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
 FT loss of appetite and tiredness.
 FT
 FT Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
 FT
 FT Query Match 100.0%; Score 25; DB 15; Length 5872;
 FT Best Local Similarity 100.0%; Pred. No. 0.0093;
 FT Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 FT
 FT QY 1 gcatctgcacccgcgcagtgacg 25
 FT ||||||||||||||||||||
 FT Db 215 gcatctgcacccgcgcagtgacg 239
 FT
 FT RESULT 5
 FT AAC74619
 FT ID AAC74619 standard; cDNA: 565 BP.
 FT XX
 FT AC AAC74619;
 FT
 FT DE Human ORFX ORF174 polynucleotide sequence SEQ ID NO:347.
 FT XX
 FT DT 08-FEB-2001 (first entry)
 FT XX
 FT XX Human: open reading frame; ORFX: detection; cytostatic; hepatotropic;
 FT KW vulnerable; antiparkinsonian; antiparkinsonian; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

XX antiviral; antibacterial; antifungal; antirheumatic; antihypertoid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 OS Homo sapiens.
 XX
 XX WO200058473-A2.
 PN
 XX
 XX 05-OCT-2000.
 PD
 XX
 XX 31-MAR-2000; 2000WO-US08621.
 PF
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 XX
 XX Shinkets RA, Leach M;
 PI
 XX WPI: 2000-602362/57.
 DR P-PSDB; AAB40410.
 DR
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PS
 XX Claim 5; Page 610; 5507pp; English.
 XX
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antirheumatic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypertoid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy.
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX Sequence 565 BP; 94 A; 168 C; 197 G; 106 T; 0 other;
 SQ

Query Match 70.4%; Score 17.6; DB 21; Length 565;
 Best Local Similarity 83.3%; Pred. No. 25;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 gcatcgatcctcgtagtga 24
 ||| |||| || |||||
 Db 390 gcaactgcgaacgcgcgtcagtga 413

RESULT 6
 AA236290/c
 ID AA236290 standard; cDNA: 241 BP.

XX
 XX AA236290;
 AC
 XX 22-FEB-2000 (first entry)
 DT
 XX Mechanical stress induced globulin-related cDNA.
 DE
 XX Expressed sequence tag; EST; mechanical stress; gene therapy;
 KW osteoporosis; bone density; bone development; ss.
 XX
 OS Rattus sp.
 XX
 XX WO9960164-A1.
 PN
 XX 25-NOV-1999.
 PD
 XX
 XX 14-MAY-1999; 99WO-US11066.
 PF
 XX
 XX 15-MAY-1998; 98US-0085673.
 PR
 XX (OUAR-) QUARK BIOTECH INC.
 PA
 XX
 XX Final P, Mor O, Skalter R, Feinstein E, Freeman A;
 PI
 XX WPI: 2000-053304/04.
 DR
 XX
 XX Identification of stress induced genes for determining risk and
 PT preventing, treating or controlling osteoporosis -
 PT
 XX
 XX Claim 25; Fig 2A-1; 308pp; English.
 XX
 XX AA236290-236320 represent genes or expressed sequence tags (ESTs)
 CC identified by the method of the invention. The sequences were isolated
 CC from rat osteoblasts. The specification describes a method for the
 CC identification of genes responsive to a specific mechanical stress. The
 CC method comprises applying the mechanical stress to an organism (tissue
 CC or cells comprising bone cells), isolating the specific cellular
 CC fractions and extracting mRNA from them, and differentially analysing the
 CC mRNA in comparison with control samples. The method is used to identify
 CC genes whose expression is responsive to a specific stress. The identified
 CC genes are employed in determining risk associated with a physiological
 CC or disease state. The risk determination methods are used for testing a
 CC medicament for gene therapy. These medicaments, or genes identified by
 CC the method of the invention, are used for treating, preventing or
 CC controlling a physiological or disease state (especially osteoporosis
 CC or bone density or other factors causing or contributing to osteoporosis
 CC or its symptoms or other conditions involved in mechanical stress or its
 CC lack. The methods can also be used for advancing research or studies in
 CC bone development.
 CC
 XX Sequence 241 BP; 51 A; 53 C; 69 G; 68 T; 0 other;
 SQ

Query Match 68.8%; Score 17.2; DB 21; Length 241;
 Best Local Similarity 86.4%; Pred. No. 35;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cacttcgacccgtagtgc 23
 ||| ||||| ||||| ||
 Db 198 cacttcgacccgtagtgc 177

RESULT 7
 AA80637/c
 ID AA80637 standard; cDNA: 393 BP.

XX AA80637;

DT 01-MAR-1999 (first entry)

XX Kidney injury associated molecule SAC_23801 cDNA clone.

XX Kidney injury associated molecule; kidney injury related molecule;

KM KIM; tissue growth promotion; regeneration; renal condition;
 XX acute renal failure; acute nephritis; tumour; ds.
 OS Rattus sp.
 XX WO9853071-A1.
 PN 26-NOV-1998.
 PD
 XX 22-MAY-1998; 98MO-US10547.
 PE
 XX 23-MAY-1997; 97US-0047491.
 PR 23-MAY-1997; 97US-0047490.
 XX
 PA (BIOJ) BIOGEN INC.
 XX
 PI Cate RL, Hession CA, Sanicola-Nadel M, Wei H;
 XX WPI; 1999-045312/04.
 DR
 XX Kidney injury-associated molecule, KIM, polypeptides - upregulated
 PT in injured or regenerating tissues, useful to promote tissue growth
 PT and regeneration, especially to treat renal conditions
 PS
 XX Claim 28; Page 191-192; 213pp; English.

CC The present sequence represents a kidney injury associated molecule
 CC (KIM) cDNA clone. KIM proteins can be administered therapeutically
 CC by expressing KIM encoding polynucleotides, to promote growth and/or
 CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins
 CC are upregulated in injured or regenerating (especially renal) tissues.
 CC KIM fusion proteins, conjugates, antibodies and vectors can also be used
 CC therapeutically, e.g. these or the KIM proteins may be included with an
 CC acceptable carrier in pharmaceutical compositions, useful for therapy/
 CC prophylaxis of conditions associated with dysfunction/dysregulation of
 CC KIM genes or proteins, especially renal diseases or impairments of renal
 CC function in humans (e.g. acute renal failure, acute nephritis). The
 CC polynucleotides can be used to produce antisense sequences which, when
 CC internalised into cells, can disrupt expression of a cellular KIM gene,
 CC also useful in therapy (e.g. to block the growth of tumours dependent on
 CC KIM for growth) or compositions. The proteins and polynucleotides are
 CC useful diagnostically e.g. to detect and quantify renal injury/disease
 CC (indicative of increased risk, or presence of, renal injury or impaired
 CC function), or abnormal responses to tissue injury (indicative of
 CC increased risk, or presence of, an autoimmune response or abnormal
 CC tissue growth arising from/affecting renal tissue). The proteins can
 CC also be used to locate KIM-producing cells (especially specific loci,
 CC e.g. tissue masses abnormally producing/expressing KIM such as tumours
 CC arising from/affecting renal tissue), by contacting cells with an
 CC imageable KIM-binding reagent and imaging reagent accumulation.
 XX
 XX
 SO Sequence 393 BP; 76 A; 108 C; 112 G; 92 T; 5 other;

Query Match 68.8%; Score 17.2; DB 20; Length 393;
 Best Local Similarity 86.4%; Pred. No. 37;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 gatttgatctctgcagtgac 23
 DB 258 CACTTCATCCTCGTCAGGGCG 237

RESULT 8
 AAC82667
 ID AAC82667 standard; cDNA; 1542 BP.
 XX
 AC AAC82667;
 XX
 DT 15-MAR-2001 (first entry)
 XX
 DE Human 7TM clone HHEN70 cDNA.
 XX

KM Transmembrane receptor protein; 7TM; gene therapy; human; bactericidal;
 KM fungicidal; viricidal; anti-HIV; analgesic; cytostatic; anabolic;
 KM anti-asthmatic; anti-parkinsonian; cardiant; hypotensive; hypertensive;
 KM osteopathic; anti-ulcer; immunosuppressive; cerebroprotective; vaccine;
 KM antidepressant; neuroprotective; hepatelical receptor; infection;
 KM serpentine receptor; G-protein coupled receptor; cancer; anorexia;
 KM bulimia; asthma; Parkinson's disease; acute heart failure; hypotension;
 KM hypertension; urinary retention; osteoporosis; angina pectoris; ulcer;
 KM myocardial infarction; allergy; benign prostatic hypertrophy; anxiety;
 KM schizophrenia; manic depression; Huntington's chorea; ss.
 XX
 XX Homo sapiens.
 OS
 XX WO200071584-A1.
 PN
 XX 30-NOV-2000.
 PD
 XX 19-MAY-2000; 2000WO-US13737.
 PE
 XX 20-MAY-1999; 99US-0135167.
 PR 13-JUL-1999; 99US-0143616.
 PR 09-SEP-1999; 99US-0152934.
 PR 14-MAR-2000; 2000US-0189029.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ruben SM, Ni J, Soppet DR, Li Y, Fan P;
 XX WPI; 2001-025139/03.
 DR P-PDB; AAB45699, AAB45706, AAB45707, AAB45708, AAB45709.
 XX
 PT Nucleic acids encoding human 7 transmembrane receptor polypeptides,
 PT useful for preventing, diagnosing and treating e.g. asthma, Parkinson's
 PT disease and ulcers -
 XX
 XX Claim 1a; Page 270; 288pp; English.

CC This invention describes novel nucleic acid molecules (I) encoding human
 CC 7 transmembrane (7TM) receptor polypeptides (also called heptalhelical,
 CC serpentine or G-protein coupled receptors). The products of the invention
 CC have bactericidal, fungicidal, viricidal, anti-HIV, analgesic,
 CC cytostatic, anabolic, anti-asthmatic, anti-parkinsonian, cardiant,
 CC hypotensive, hypertensive, osteopathic, anti-ulcer, immunosuppressive,
 CC cerebroprotective, antidepressant and neuroprotective activity and can
 CC be used in gene therapy or in a vaccine. (I) and the 7TM receptor protein
 CC it encodes may be used in the prevention, treatment and diagnosis of
 CC diseases associated with inappropriate 7TM receptor expression. (I) and
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays (e.g. polymerase chain reactions (PCR)) to detect and quantitate
 CC the presence of similar nucleic acid sequences in samples, and hence
 CC which patients may be in need of restorative therapy. Diseases associated
 CC with inappropriate 7TM receptor polypeptide expression include for
 CC example bacterial, fungal and viral (especially human immuno-deficiency
 CC virus) infections, pain, cancers, anorexia, bulimia, asthma, Parkinson's
 CC disease, acute heart failure, hypotension, hypertension, urinary
 CC retention, osteoporosis, angina pectoris, myocardial infarction, ulcers,
 CC allergies, benign prostatic hypertrophy, anxiety, schizophrenia, manic
 CC depression and Huntington's chorea.
 XX
 XX

SO Sequence 1542 BP; 451 A; 325 C; 306 G; 460 T; 0 other;

Query Match 68.0%; Score 17; DB 22; Length 1542;
 Best Local Similarity 80.0%; Pred. No. 55;
 Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 gcatctcttctcaccagtgacg 25
 DB 1451 gcatctcttctcaccagtgacg 1475

RESULT 9
 AAF33079


```

RESULT 11
AAQ15019/C
ID AAQ15019 standard; DNA: 3234 BP.
XX
AC AAQ15019;
XX
XX 25-MAR-1992 (first entry)
XX
DE Encodes mouse bullous pemphigoid antigen protein.
XX
KM dermatological syndrome; blisters; murine; ss.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 1..2894
FT /tag= a
XX
XX JP03206885-A.
XX
XX 10-SEP-1991.
XX
XX 09-APR-1990; 90JP-0093585.
XX
XX 31-OCT-1989; 89JP-0283947.
XX
XX 09-APR-1990; 90JP-0093585.
XX
XX (TOFU) TONEN CORP.
XX
XX WPI: 1991-356566/49.
XX
XX P-PSDB; AAR15345.
XX
XX PT DNA which encodes mouse derived bullous pemphigoid antigen protein -
XX contg. base sequence which encodes polypeptide composed of specified
XX 996 aminoacid sequence
XX
XX PS Claim 1; Fig 1; 23pp; Japanese.
XX
XX CC The polypeptide encoded by this nucleotide sequence is recognised
XX by auto-antibodies in the serum of a bullous pemphigoid patient and
XX by a human anti-basement membrane zone monoclonal antibody.
XX
XX CC See also AAQ15020.
XX
XX SO Sequence 3234 BP; 938 A; 738 C; 873 G; 685 T; 0 other;

Query Match 68.0%; Score 17; DB 12; Length 3234;
Best Local Similarity 80.0%; Pred. No. 61;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcatcgcacccgcgcaggtcag 25
   ||||||| |||||||
DB 235 GCCTTCGATGCTTCAGGTGCG 211

RESULT 12
AAZ30163/C
ID AAZ30163 standard; DNA: 34094 BP.
XX
AC AAZ30163;
XX
XX 26-JAN-2000 (first entry)
XX
DE Complete nucleotide sequence of the PAV-3 genome.
XX
XX PAV-3; defective recombinant PAV vector; live recombinant virus;
XX subunit vaccine; nucleic acid immunisation; gene therapy;
XX genetic disease; hemophilia; cystic fibrosis; cancer; viral infection;
XX acquired immune deficiency syndrome; PAV antigen; porcine pathogen; ds.
XX
XX OS Porcine adenovirus Type 3.
XX
XX PN MO953047-A2.

```

```

XX
XX 21-OCT-1999.
XX
XX 15-APR-1999; 99WO-US08220.
XX
XX 15-APR-1998; 98US-0081882.
XX
XX (UYSA-) UNIV SASKATCHEWAN.
XX
XX Reddy PS, Tikoo SK, Babluk LA;
XX
XX WPI: 1999-620422/53.
XX
XX New nucleic acids from the genome of porcine adenovirus-3, and derived
XX gene therapy vectors, particularly for immunization
XX
XX Example 2; Fig 1; 87pp; English.
XX
XX The present sequence represents the complete nucleotide sequence of the
XX genome of porcine adenovirus-3 (PAV-3). The specification also describes
XX a defective recombinant PAV vector comprising inverted terminal repeats
XX (ITR), packaging sequences and at least one heterologous nucleotide
XX sequence (II), but lacking E1 functions. The defective vectors replicate
XX inefficiently in cells (other than helper cells) so are unlikely to be
XX immunogenic. Deletion of the E1 (and optionally other regions) increases
XX the size of heterologous insert that can be packaged. The PAV-3
XX polynucleotides sequences are used to produce (recombinant or defective)
XX vectors that can express heterologous proteins, e.g. for making live,
XX recombinant virus or subunit vaccines, for nucleic acid immunisation or
XX for gene therapy (e.g. of genetic diseases such as hemophilia or cystic
XX fibrosis, cancer, or viral infections, including acquired immune
XX deficiency syndrome), also for in vitro expression of recombinant
XX antigens (for antibody production), antisense RNA, ribozymes or
XX therapeutic proteins. They are also used diagnostically to detect PAV
XX antigens and/or nucleic acid. The vectors may be used in human or
XX veterinary medicine, but particularly for expressing protective
XX determinants of porcine pathogens. Regulatory regions may be used to
XX control expression of heterologous genes. Antibodies raised against PAV-3
XX polypeptides can also be used for diagnosis (to detect PAV-specific
XX antigen).
XX
XX SO Sequence 34094 BP; 6240 A; 11070 C; 10693 G; 6091 T; 0 other;

Query Match 66.4%; Score 16.6; DB 20; Length 34094;
Best Local Similarity 82.6%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 attcgcacccgcgcaggtcag 25
   ||||||| |||||||
DB 27375 ATTTCGTTACGCGCAGGTGCG 27353

RESULT 13
AAC56248/C
ID AAC56248 standard; DNA: 385 BP.
XX
AC AAC56248;
XX
XX 25-JAN-2001 (first entry)
XX
DE Pinus radiata transcription factor DNA sequence #48.
XX
XX Pinus radiata transcription factor; gene expression; eucalyptus; pine; acacia;
XX poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
XX basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
XX homeodomain zipper; LIM domain; AP2; ERBBs; zinc finger domain;
XX type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
XX OS Pinus radiata.
XX
XX PN WO200053724-A2.

```

PD 14-SEP-2000.
 XX
 PF 09-MAR-2000; 2000WO-US06112.
 XX
 PR 11-MAR-1999; 99US-0266513.
 XX 18-AUG-1999; 99US-0149485.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Wood M, McGrath A, Shenk MA, Glenn M;
 XX
 DR WPI; 2000-579369/54.
 XX
 PT New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT
 PS Claim 1; Page 143; 747pp; English.
 XX
 CC The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.
 XX
 SQ Sequence 385 BP; 128 A; 77 C; 94 G; 86 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 385;
 Best Local Similarity 85.7%; Pred. No. 1.1e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 ttgcgacccgcgcaggtgca 24
 |||||
 Db 31 TTTCGATCTCTCTCAGCTGGA 11

RESULT 14
 AAC57313/C
 ID AAC57313 standard; DNA; 393 BP.
 XX
 AC AAC57313;
 XX
 DT 25-JAN-2001 (first entry)
 XX
 DE Pinus radiata transcription factor DNA sequence #674.
 XX
 KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW Poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.
 XX
 OS Pinus radiata.
 XX
 PN WO200053724-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 09-MAR-2000; 2000WO-US06112.
 XX
 PR 11-MAR-1999; 99US-0266513.
 XX 18-AUG-1999; 99US-0149485.
 XX

PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Wood M, McGrath A, Shenk MA, Glenn M;
 XX
 DR WPI; 2000-579369/54.
 XX
 PT New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT
 PS Claim 1; Page 643; 747pp; English.
 XX
 CC The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and ERBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.
 XX
 SQ Sequence 393 BP; 133 A; 78 C; 95 G; 87 T; 0 other;

Query Match 64.8%; Score 16.2; DB 21; Length 393;
 Best Local Similarity 85.7%; Pred. No. 1.1e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 ttgcgacccgcgcaggtgca 24
 |||||
 Db 31 TTTCGATCTCTCTCAGCTGGA 11

RESULT 15
 AAX30823
 ID AAX30823 standard; DNA; 865 BP.
 XX
 AC AAX30823;
 XX
 DT 20-MAY-1999 (first entry)
 XX
 DE Streptococcus pneumoniae genomic DNA sequence SEQ ID NO:100.
 XX
 KW Streptococcus pneumoniae strain 0100993; vaccine; immune response;
 KW streptococcal infection; pneumococcal; ss.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO9737026-A1.
 XX
 PD 09-OCT-1997.
 XX
 PF 01-APR-1997; 97WO-US05306.
 XX
 PR 22-AUG-1996; 96US-0025788.
 XX 02-APR-1996; 96US-0014690.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 XX
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 XX Stodola RK;
 XX
 DR WPI; 1997-503111/46.
 XX P-PSDB; AAY11240.
 XX
 PT Nucleic acids encoding pneumococcal polypeptide(s) - useful in

PT vaccines, drug screening, etc
 XX
 PS Claim 5; Page 119; 354pp; English.
 XX
 CC AAX30722 to AAX30946 represent genomic DNA sequences isolated from
 CC Streptococcus pneumoniae strain 0100993. These genomic DNA sequences
 CC encode the novel proteins given in AAT1114 to AAT11367. The proteins,
 CC isolated from Streptococcus pneumoniae, can be used in vaccines against
 CC streptococcal infections and in assays for identifying compounds that
 CC inhibit or activate the activity of the proteins. The antagonists can
 CC be used to treat an individual having need to inhibit a bacterial
 CC protein. Vectors expressing the proteins can be used to induce a
 CC protective immune response in mammals.
 XX
 XX Sequence 865 BP; 270 A; 215 C; 157 G; 223 T; 0 other;
 S0

Query Match 64.8%; Score 16.2; DB 18; Length 865;
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 attcgatcctgcaggtgc 23
 ||||| ||||| ||||| |||||
 Db 115 attccagcctgtcaggtgc 135

RESULT 16
 AAT98665/C
 ID AAT98665 standard; DNA; 882 BP.
 XX
 AC AAT98665;
 XX
 DT 09-NOV-1998 (first entry)
 XX
 DE DNA encoding a S. pneumoniae protein of unknown function.
 XX
 KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;
 KW immunological response; inoculation; antibody production; inhibitor;
 KW T cell immune response; antimicrobial compound; bacterial adhesion;
 KW extracellular matrix protein; protein-mediated cell invasion; wound;
 KW pathogenesis; ss.
 XX
 OS Streptococcus pneumoniae.
 XX
 FH Key Location/Qualifiers
 FT CDS 120..503
 FT /*tag= a
 XX
 XX WO9743303-A1.
 XX
 PD 20-NOV-1997.
 XX
 PF 14-MAY-1997; 97WO-US07950.
 XX
 PR 14-MAY-1996; 96US-0017670.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA (SMIK) SMITHKLINE BEECHAM PLC.
 XX
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO;
 PI Stodola RK;
 XX
 DR WPI: 1998-008793/01.
 DR P-PSDB; AAW38617.
 XX
 XX Novel Streptococcus pneumoniae proteins and related DNA - useful for
 PT diagnosing anti-microbial agents for treatment of bacterial
 PT infections
 XX
 PS Claim 4; Page 182; 483pp; English.
 XX
 CC This sequence encodes a Streptococcus pneumoniae protein of unknown
 CC function, and represents a DNA sequence of the invention.

CC The DNA sequences were isolated from Streptococcus pneumoniae strain
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
 CC invention can be used to identify compounds which interact with and
 CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of
 CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.
 XX
 XX Sequence 882 BP; 229 A; 162 C; 213 G; 278 T; 0 other;
 S0

Query Match 64.8%; Score 16.2; DB 19; Length 882;
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 attcgatcctgcaggtgc 23
 ||||| ||||| ||||| |||||
 Db 768 ATTTCAGCCTGTCTCAGTGC 748

RESULT 17
 AA296336
 ID AA296336 standard; DNA; 1702 BP.
 XX
 AC AA296336;
 XX
 DT 10-APR-2000 (first entry)
 XX
 DE S. pneumoniae derived DNA from ORF #164.
 XX
 KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
 KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO9806734-A1.
 XX
 PD 19-FEB-1998.
 XX
 PF 15-AUG-1997; 97WO-US14436.
 XX
 PR 16-AUG-1996; 96US-0024022.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PA Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
 PI Stodola RK;
 XX
 DR WPI: 1998-159452/14.
 DR P-PSDB; AAY85994, AAY85995.
 XX
 XX Streptococcus pneumoniae proteins and related DNA - useful for
 PT screening compounds for antibacterial activity
 PT
 XX
 PS Claim 4; Page 198-199; 640pp; English.
 XX
 CC This invention describes novel isolated Streptococcus pneumoniae
 CC polynucleotides (see AA296173-296494) and their encoded proteins (see
 CC AA85792-Y86182). The DNA, vectors and host cells described in the
 CC method of the invention are useful for the recombinant expression of the
 CC polypeptides. The polypeptides are useful for treatment or prevention of
 CC disease, or diagnosis of disease related to expression or activity of
 CC such a polypeptide. They can also be used to screen for compounds which
 CC interact with and inhibit or activate such a polypeptide. The

CC polypeptides (or DNA encoding them, via gene therapy) are also useful
 CC for inducing an immunological response in a mammal. The antagonists are
 CC useful to inhibit such bacterial polypeptides. The polypeptides are
 CC particularly useful to identify antimicrobial compounds and antibiotics.
 CC They are also useful to determine their role in pathogenesis of
 CC infection, dysfunction and disease.

XX Sequence 1702 BP; 490 A; 397 C; 348 G; 467 T; 0 other;

Query Match 64.8%; Score 16.2; DB 19; Length 1702;
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 attcgatccctgcagtgac 23
 ||||| | ||| |||||
 Db 899 attccagcctgtcagtgac 919

RESULT 18

AAV37384
 ID AAV37384 standard; DNA; 1702 BP.

XX AAV37384;

DT 13-OCT-1998 (first entry)

DE Streptococcus pneumoniae coding region.

XX coding region; ORF; Open reading frame; antibacterial;
 XX infection; prevention; meningitis; ss.

OS Streptococcus pneumoniae.

XX Key Location/Qualifiers
 FT CDS complement (1164..1640)
 FT /*tag= a
 FT /product= putative AGAS protein

XX W09819689-A1.

XX 14-MAY-1998.

XX 27-OCT-1997; 97WO-US19226.

XX 01-NOV-1996; 96US-0029930.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (SMIK) SMITHKLINE BEECHAM PLC.

XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
 XX Reid RH, Zarfos PN;

XX WPI: 1998-286586/25.

XX P-PSDB; AAM60994.

XX New isolated nucleic acids from Streptococcus pneumoniae - useful,
 XX e.g. for identifying anti-bacterial(s) for treatment and prevention
 XX of meningitis

XX Claim 1; Page 89-90; 130pp; English.

XX The sequence is that of a coding region isolated from
 CC S. pneumoniae. Its encoded protein, or agonists of it,
 CC may be useful as an antibacterial for treatment or
 CC prevention of infection, specifically caused by S. pneumoniae
 CC (particularly meningitis) but possibly also Helicobacter
 CC pylori (ulcers and gastric cancer). It may be of particular
 CC use before insertion of an in-dwelling device or any other
 CC invasive procedure. The protein, or nucleic acid encoding
 CC it, can also be used in vaccines to induce a cellular
 CC and/or humoral immune response, or to screen for other
 CC antibacterials. The DNA may also contain flanking sequences

CC that are potential sources of control elements for bacterial
 CC gene expression. Detecting a sequence encoding the protein
 CC can be used diagnostically, e.g. to detect a mutation for
 CC serotyping or classifying infectious agents.

XX Sequence 1702 BP; 490 A; 397 C; 348 G; 467 T; 0 other;

Query Match 64.8%; Score 16.2; DB 19; Length 1702;
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 attcgatccctgcagtgac 23
 ||||| | ||| |||||
 Db 899 attccagcctgtcagtgac 919

RESULT 19

AAV52304/c
 ID AAV52304 standard; DNA; 14736 BP.

XX AAV52304;

DT 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:171.

XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 XX computer readable medium; vaccine; pharmaceutical composition; ds.

OS Streptococcus pneumoniae.

XX W09818931-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US19588.

XX 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 XX Kunsch CA, Rosen CA;

XX WPI: 1998-272225/24.

XX Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PT pneumoniae

XX Claim 1; Page 1085-1094; 1409pp; English.

XX The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment of a sequence at least 95%
 CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridise to the target sequence and
 CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC molecules whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or

CC expression modulating fragments of the *S. pneumoniae* genome. Products
 CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for *S. pneumoniae*.

XX Sequence 14736 BP; 4665 A; 2568 C; 3255 G; 4247 T; 1 other;

Query Match 64.8%; Score 16.2; DB 19; Length 14736;

Best Local Similarity 85.7%; Pred. No. 1.8e+02; Mismatches 3; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db 6166 ATTTCACGCTTGTCAGCTGC 6146

0Y 3 attcgcctcgcgcagtcag 23
 ||||| ||||| ||||| |||||

RESULT 20
 AAT59977/C
 ID AAT59977 standard; DNA; 1045 BP.

XX AAT59977;

DT 12-MAY-1997 (first entry)

XX 5' untranslated region of the *tipe* protein coding sequence.

XX *Drosophila*; *tipe*; para protein; voltage-dependent cation channel; stroke;
 KW pesticide; insecticide; insect; parasitic infection; human; head trauma;
 KW neuroprotection; hypoxia; therapy; ss.

XX *Drosophila melanogaster*.

OS US5593862-A.

PN 14-JAN-1997.

PD 04-OCT-1994; 94US-0317880.

PF 04-OCT-1994; 94US-0317880.

PR (UYNV) UNIV NEW YORK STATE RES FOUND.

PA Feng G, Hall LM;

PI WPI; 1997-099467/09.

XX Nucleic acid encoding *Drosophila melanogaster tipe* protein - for
 PT prodn. of recombinant voltage-dependent cation channel

XX Example 11; Column 35-36; 33pp; English.

XX AAT59977 and AAT59978 represent the 5' and 3' untranslated regions of
 CC the *Drosophila tipe* protein (see AAM1343) coding sequence (see
 CC AAT59975). Mutations in the *tipe* protein, result in a ethyl methane
 CC sulphoxide- induced recessive mutation phenotype. Homologous flies for
 CC the mutation paralyze rapidly at 38 degrees, and recover immediately when
 CC returned to 23 degrees. Coexpression of the full length *tipe* sequence,
 CC and a nucleic acid encoding a para protein results in translation
 CC products that form a functional voltage-dependent cation channel. The
 CC cation channel can be used to screen for pesticides active against
 CC insects such as *Drosophila melanogaster* and pest insects. The cation
 CC channel can also be used to screen for drugs for use in the treatment and
 CC prevention of parasitic infections in humans and animals, and to screen
 CC drugs for their neuroprotective effect against hypoxia, stroke, and head
 CC trauma.

XX Sequence 1045 BP; 379 A; 237 C; 242 G; 187 T; 0 other;

Query Match 64.0%; Score 16; DB 18; Length 1045;
 Best Local Similarity 79.2%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

0Y 2 cattcgcctcgcgcagtcag 25
 ||||| ||||| ||||| |||||

Db 163 CCTTCTATCTTGCCACGTCAG 140

RESULT 21
 AAX02016/C
 ID AAX02016 standard; DNA; 1045 BP.

XX AAX02016;

DT 21-APR-1999 (first entry)

DE D. *melanogaster tipe+ 4kb* clone 5'-UTR.

XX *tipe+*; para protein; modulator; voltage dependent cation channel; VDCC;
 KW pesticide; insect control; pharmaceutical agent; neuroprotection;
 KW hypoxia; ischaemia; stroke; head trauma; ss.

XX *Drosophila melanogaster*.

OS Key Location/Qualifiers

FT 5'UTR 1..1045

FT /*tag= a

PN US5871940-A.

PD 16-FEB-1999.

PF 13-JAN-1997; 97US-0782396.

PR 04-OCT-1994; 94US-0317880.

PR 13-JAN-1997; 97US-0782396.

PA (UYNV) UNIV NEW YORK STATE RES FOUND.

PI Feng G, Hall LM;

PI WPI; 1999-16632/14.

XX Screening for agents which modulate ion channel function - using
 PT host cells transformed with nucleic acids encoding the *Drosophila*
 PT *melanogaster tipe* and para proteins

PS Disclosure; Column 35-36; 54pp; English.

XX This sequence represents the 5'-UTR from a *Drosophila melanogaster tipe+*
 CC protein which is used in a method for screening for agents which modulate
 CC ion channel function which uses host cells transformed with nucleic acid
 CC encoding *Drosophila melanogaster tipe* and para proteins. Co-expression of
 CC these genes in the host cell, allows the formation of a functional
 CC voltage dependent cation channel (VDCC) in the cell. The agents
 CC identified can be used as pesticides for the control of *Drosophila*
 CC *melanogaster* or other insects. They can also be used to screen
 CC pharmaceutical agents for their neuroprotective affect against e.g.
 CC hypoxia, ischaemia, stroke and head trauma.

XX Sequence 1045 BP; 379 A; 237 C; 242 G; 187 T; 0 other;

Query Match 64.0%; Score 16; DB 20; Length 1045;
 Best Local Similarity 79.2%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

0Y 2 cattcgcctcgcgcagtcag 25
 ||||| ||||| ||||| |||||

Db 163 CCTTCTATCTTGCCACGTCAG 140

RESULT 22
 AAX36567/C
 ID AAX36567 standard; DNA; 1137 BP.

```

AC  AAX36567;
XX
DT  07-JUL-1999 (first entry)
XX
DE  Human EDG-3 coding sequence.
XX
KW  EDG-1; EDG-2; EDG-3; EDG-4; EDG-5; PSP-24; human; detection; therapy;
KM  inverse agonist; allosteric modulator; lysophosphatidic acid receptor;
XX  LPA signalling mediated disease; cellular apoptosis; ss.
OS  Homo sapiens.
XX
PN  W09919513-A2.
XX
PD  22-APR-1999.
XX
PF  09-OCT-1998; 98WO-US21315.
XX
PR  10-OCT-1997; 97US-0061572.
XX
PA  (LXRB-) LXR BIOTECHNOLOGY INC.
XX
PI  Erikson J, Goddard JG, Kiefer M;
XX  WPI: 1999-277658/23.
DR  P-PSDB; AAI05490.
XX
PT  Identification of (ant)agonists of LPA receptor EDG-2 for, e.g.
XX  treating LPA signalling mediated diseases such as cellular apoptosis
XX
PS  Claim 1; Page 53-54; 63pp; English.
XX
CC  This sequence encodes human EDG-3.
CC  The invention relates to methods of detecting (ant)agonist, inverse
CC  agonist or allosteric modulators of the lysophosphatidic acid receptors
CC  EDG-1, EDG-2, EDG-3, EDG-4, EDG-5, and PSP-24. The methods are used to
CC  identify (ant)agonists and allosteric modulator of the lysophosphatidic
CC  acid (LPA) EDG2 receptor, e.g. to treat LPA signalling mediated disease
CC  such cellular apoptosis.
XX
SQ  Sequence 1137 BP; 226 A; 362 C; 298 G; 251 T; 0 other;

Query Match          64.0%; Score 16; DB 20; Length 1137;
Best Local Similarity 79.2%; Pred. NO. 1.6e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY  2 cattcgatcctgcagtgacg 25
    ||||| |||| | |||| |||| |
DB  429 CATTTGATCATGTGTCAAGTCCG 406

RESULT 23
AAZ22300/C
ID  AAZ22300 standard; cDNA; 1137 BP.
XX
AC  AAZ22300;
XX
DT  28-MAR-2000 (first entry)
XX
DE  Human EDG3sb polypeptide encoding cDNA.
XX
KW  EDG3sb protein; infection; HIV-1; HIV-2; pain; cancer; diabetes; obesity;
KM  anorexia; bulimia; asthma; Parkinson's disease; acute heart failure;
XX  hypotension; hypertension; urinary retention; osteoporosis; allergy;
KM  angina pectoris; myocardial infarction; stroke; ulcer; psychotic; human;
XX  neurological disorder; dyskinesia; Huntington's disease; antidepressant;
KM  neurotropic; antimigraine; antiemetic; ss.
XX
OS  Homo sapiens.
XX
XX  Key Location/Qualifiers
FH  1..1137
FT  CDS

```

```

FT  /tag= a
FT  /transl_except= (pos:568..570, aa: Pro)
FT  /product= "EDG3sb polypeptide"
XX
PN  W09960019-A1.
XX
PD  25-NOV-1999.
XX
PF  12-MAY-1999; 99WO-US10337.
XX
PR  20-MAY-1998; 98US-0082088.
XX
PA  (SMK ) SMITHKLINE BEECHAM CORP.
XX
PI  Tsui P;
XX
DR  WPI: 2000-116310/10.
XX  P-PSDB; AAI59176.
XX
PT  Novel polypeptides, used to develop products for treating e.g.
XX  infections, cancers, diabetes, obesity, asthma, Parkinson's disease,
XX  heart failure, stroke or neurological disorders
XX
PS  Claim 13; Page 27; 36pp; English.
XX
CC  This cDNA encodes a human EDG3sb polypeptide. The EDG3sb protein can be
CC  expressed by standard recombinant methodology. The products can be used
CC  for treating e.g. infections such as bacterial, fungal, protozoan and
CC  viral infections, particularly infections caused by HIV-1 or HIV-2, pain,
CC  cancers, diabetes, obesity, anorexia, bulimia, asthma, Parkinson's
CC  disease, acute heart failure, hypotension, hypertension, urinary
CC  retention, osteoporosis, angina pectoris, myocardial infarction, stroke,
CC  ulcers, allergies, benign prostatic hypertrophy, migraine, vomiting,
CC  psychotic and neurological disorders, including anxiety, schizophrenia,
CC  manic depression, depression, delirium, dementia, and severe mental
CC  retardation, and dyskinesias, such as Huntington's disease or Gilles de
CC  Tourette's syndrome, related to either an excess of, or an under-
CC  expression of EDG3sb polypeptide activity. The products can also be used
CC  for detection, diagnosis and drug screening.
XX
SQ  Sequence 1137 BP; 225 A; 362 C; 299 G; 251 T; 0 other;

Query Match          64.0%; Score 16; DB 21; Length 1137;
Best Local Similarity 79.2%; Pred. NO. 1.6e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY  2 cattcgatcctgcagtgacg 25
    ||||| |||| | |||| |||| |
DB  429 CATTTGATCATGTGTCAAGTCCG 406

RESULT 24
AAA65913/C
ID  AAA65913 standard; DNA; 1152 BP.
XX
AC  AAA65913;
XX
DT  05-OCT-2000 (first entry)
XX
DE  E. coli proliferation associated coding sequence SEQ ID NO:105.
XX
KW  Escherichia coli; E. coli; proliferation; inhibition; screening;
KM  antimicrobial; bacterial growth; antisense therapy; antibacterial; ds.
XX
OS  Escherichia coli.
XX
PN  W0200044906-A2.
XX
PD  03-AUG-2000.
XX
PF  27-JAN-2000; 2000WO-US02200.
XX

```

PR 27-JAN-1999; 9905-0117405.
 XX (ELIT-) ELITRA PHARM INC.
 XX Zykind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ,
 PI Yamamoto RT, Xu HH;
 XX WPI: 2000-514822/46.
 DR P-PSDB: AAB15908.
 XX
 PT Novel polynucleotides and polypeptides associated with microorganism
 PT proliferation, used to identify inhibitors of bacterial growth and
 PT proliferation, for use in antisense therapy -
 PS
 PS Claim 8; Page 123; 316pp; English.
 XX
 XX AA65809 to AA65889 and AA66058 to AA66138 represent nucleotide
 CC sequences derived from *Escherichia coli* which inhibit *E. coli*
 CC proliferation. AA65890 to AA66055 and AAB15886 to AAB16040 represent
 CC nucleotide and protein sequences associated with *E. coli* proliferation.
 CC AA66056 and AA66057 represent primers used for sequencing *E. coli*
 CC proliferation inhibiting nucleotide inserts in an example from the
 CC present invention. Methods from the present invention can be used to
 CC identify a proliferation-required gene in a microorganism, by contacting
 CC a microorganism with a proliferation-required gene activity inhibitory
 CC nucleic acid identified in another organism, and determining if
 CC inhibition occurs in the second microorganism. The nucleic acid sequences
 CC identified as being required for bacterial growth and proliferation, can
 CC be used for antisense therapy for killing bacteria.
 XX
 SQ Sequence 1152 BP; 267 A; 270 C; 354 G; 261 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 1152;
 Best Local Similarity 79.2%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 gcattcgatcctgcagtgca 24
 |||| ||||||| |||| ||||
 DB 260 GCATCGCATCTCATCAGCGCA 237

RESULT 25
 AAA39397/c
 ID AAA39397 standard; DNA: 1152 BP.
 XX
 AC AAA39397;
 XX
 DT 15-SEP-2000 (first entry)
 XX
 DE *Escherichia coli* monomeric FtsZ nucleotide sequence SEQ ID NO:5.
 XX
 KW *Staphylococcus aureus*; monomeric FtsZ; monomeric FtsA; antibacterial;
 KW *Escherichia coli*; multimeric protein; bacterial infection; vaccine;
 KW microbial infection; screening; ds.
 XX
 OS *Escherichia coli*.
 XX
 PN W0200024252-A1.
 XX
 PD 04-MAY-2000.
 XX
 PF 20-OCT-1999; 99WO-US24653.
 XX
 PR 23-OCT-1998; 98US-0105315.
 PR 19-APR-1999; 99US-0129965.
 PR 03-MAY-1999; 99US-0132333.
 XX
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 XX
 PI Yan K, Pearce KH;
 XX
 DR WPI: 2000-350517/30.

DR P-PSDB: AAY91110.
 XX
 PT Multimeric FtsZ:FtsA polypeptide, useful as a vaccine for treating
 PT bacterial infections and for diagnosing diseases associated with
 PT bacterial infection -
 XX
 XX
 PS Disclosure; Page 4; 45pp; English.
 XX

CC The present invention describes isolated multimeric FtsZ:FtsA
 CC polypeptides (1) comprising the protein sequences given in AAY91108,
 CC AAY91109, AAY91110 and AAY91111. Also described are: (1) an antibody
 CC (Ab) specific for (1); (2) a method (M1) for treating an individual
 CC requiring enhanced or reduced activity or expression of (1) comprising
 CC administering an agonist or antagonist to (1), respectively; (3) a
 CC method (M2) for diagnosing a disease or susceptibility to a disease
 CC related to expression or activity of (1) comprising determining the
 CC presence or absence of a mutation in the nucleotide sequence encoding
 CC (1) and/or analysing for the presence or quantity of (1) in a sample;
 CC and (4) a method (M3) for screening to identify compounds that activate
 CC or inhibit the function of (1) selected from: (a) measuring the binding
 CC of a candidate compound to the polypeptide using a label; (b) measuring
 CC the binding of a candidate compound to (1) in the presence of a labeled
 CC competitor; (c) testing whether the candidate compound results in a
 CC signal generated by activation or inhibition of (1); or (d) detecting
 CC the effect of a candidate compound on the production of mRNA encoding
 CC (1) using e.g. enzyme linked immunosorbent assay (ELISA). AAY91108 to
 CC AAY91111 are useful as vaccines for treating microbial infections such
 CC as *Staphylococcus*, *Streptococcus* and *Escherichia coli* infections. The
 CC polypeptides are also useful for screening for antibacterial compounds
 CC and for detecting diseases associated with microbial infections. The
 CC present sequence encodes *Escherichia coli* FtsZ.
 XX

SQ Sequence 1152 BP; 267 A; 270 C; 354 G; 261 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 1152;
 Best Local Similarity 79.2%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 gcattcgatcctgcagtgca 24
 |||| ||||||| |||| ||||
 DB 260 GCATCGCATCTCATCAGCGCA 237

RESULT 26
 AAX59380/c
 ID AAX59380 standard; DNA: 1792 BP.
 XX
 AC AAX59380;
 XX
 DT 20-SEP-1999 (first entry)
 XX
 DE *Aspergillus nidulans* essential gene AN85.
 XX
 KW AN85; essential gene; fungicide; fungistatic; screening; infection;
 KW aspergillosis; therapy; ds.
 XX
 OS *Aspergillus nidulans*.
 XX
 PN Aspergillus nidulans.
 XX
 FH Key Location/Qualifiers
 FT CDS 230..1478
 FT
 FT /tag= a
 FT /note= "contains introns"
 FT /note= 309
 FT exon
 FT /tag= b
 FT /number= 1
 FT 310..374
 FT intron
 FT /tag= c
 FT /number= 1
 FT exon
 FT /tag= d
 FT /number= 2
 FT 819..875
 FT intron

Y-chromosome specific nucleic acid sequences - used for determining the sex of bovine or ruminant embryo, fetus or sex chromosome content of sperm or other cells

Disclousure: Page 34, 51pp; English.

The sequences given in AA023962 and AA023963 are both sensitive Y-chromosome probes, designated Bty1 and Bty2 respectively. Bty1 is a 1.859 kb PstI fragment whereas Bty2 is a 3.71 kb SacI fragment, both of which were cloned from bovine DNA. Using these probes it is possible to sex ruminant/bovine embryos by nucleic acid hybridisation, rapidly and with virtually 100% accuracy. The amount of DNA required is equal to the amount obid. from two or fewer embryonic cells. The probes can be used for determining the sex of an embryo, fetus or the sex chromosome content of sperm or other cells. The probes can also be used to detect variations in amts. and/or variations in sequence of corresponding sequences in individual mammals, eg. in paternity testing of male offspring.

Sequence 3708 BP; 883 A; 1053 C; 798 G; 973 T; 1 other;

Query Match 64.0%; Score 16; DB 13; Length 3708;

Best Local Similarity 79.2%; Pred. No. 1.9e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 gcattcgatccctgcaggtgca 24

1533 GCAGATCGCTCCTGTGAGAGCA 1510

RESULT 29

AAA53179/C

ID AAA53179 standard; DNA; 3904 BP.

AAA53179;

06-OCT-2000 (first entry)

Streptococcus bovis plasmid PSB01 nucleotide sequence SEQ ID NO:1.

Streptococcus bovis; autonomously replicating sequence; ARS;

Colibacillus; shuttle vector; lumen microorganism; ds.

Streptococcus bovis.

JP304303-B1.

22-MAY-2000.

07-DEC-1998; 98JP-0347017.

07-DEC-1998; 98JP-0347017.

(NORO) NORINSUISANSHO CHIKUSAN SHIKENJOCHO.

WPI; 2000-415207/36.

Plasmid used as a shuttle vector between Streptococcus bovis and Colibacillus contains autonomously replicating sequence with specified base pairs

Claim 2; Page 5-6; 9pp; Japanese.

The present invention describes a plasmid used as a shuttle vector between Streptococcus bovis and Colibacillus, which contains an autonomously replicating sequence (ARS) of S. bovis origin with a polynucleotide sequence of 3904 base pairs (the present sequence), or the DNA which is hybridised under stringent conditions to the polynucleotide sequence. The plasmid acts as a shuttle vector between Streptococcus bovis and Colibacillus.

Sequence 3904 BP; 1447 A; 554 C; 743 G; 1160 T; 0 other;

Query Match 64.0%; Score 16; DB 21; Length 3904;

Best Local Similarity 79.2%; Pred. No. 1.9e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 gcattcgatccctgcaggtgca 24

2721 GCTTTAAATCTTCGTTAGTGCA 2698

RESULT 30

AAT5975/C

ID AAT5975 standard; DNA; 3954 BP.

AAT5975;

12-MAY-1997 (first entry)

Full length tpe protein coding sequence.

Drosophila; tpe; para protein; voltage-dependent cation channel; stroke;

pesticide; insecticide; insect; parasitic infection; human; head trauma;

neuroprotection; hypoxia; therapy; ss.

Drosophila melanogaster.

Key Location/Qualifiers

CDS 1046..2404

/*tag= a

/product= tpe

US5593862-A.

14-JAN-1997.

04-OCT-1994; 94US-0317880.

04-OCT-1994; 94US-0317880.

(UYNV) UNIV NEW YORK STATE RES FOUND.

Feng G, Hall LM;

WPI; 1997-099467/09.

P-PSDB; AAM13843.

Nucleic acid encoding Drosophila melanogaster tpe protein - for

prodn. of recombinant voltage-dependent cation channel

Disclousure; Column 29-32; 33pp; English.

This sequence represents the full length coding sequence for the Drosophila tpe protein. Mutations in the tpe protein, result in a ethyl methane sulphonate-induced recessive mutation phenotype. Homologous flies for the mutation paralyse rapidly at 38 degrees, and recover immediately when returned to 23 degrees. Coexpression of this sequence, and a nucleic acid encoding a para protein results in translation products that form a functional voltage-dependent cation channel. The cation channel can be used to screen for pesticides active against insects such as Drosophila melanogaster and pest insects. The cation channel can also be used to screen for drugs for use in the treatment and prevention of parasitic infections in humans and animals, and to screen drugs for their neuroprotective effect against hypoxia, stroke, and head trauma.

Sequence 3954 BP; 1293 A; 960 C; 867 G; 834 T; 0 other;

Query Match 64.0%; Score 16; DB 18; Length 3954;

Best Local Similarity 79.2%; Pred. No. 1.9e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cattgatcctcgtcaggtgcag 25
 ||||| ||||| ||||| |||||
 DB 163 CCTTCTATCCTTGCACGTCGAG 140

RESULT 31

AAK02015/c
 ID AAK02015 standard; DNA: 3954 BP.

AAK02015;

DT 21-APR-1999 (first entry)

DE D. melanogaster t1pe+ genomic DNA.

KW t1pe+; para protein; modulator; voltage dependent cation channel; VDCC;

KM pesticide; insect control; pharmaceutical agent; neuroprotection;

KW hypoxia; ischaemia; stroke; head trauma; ss.

OS Drosophila melanogaster.

XX Key Location/Qualifiers

FT CDS 1046..2404

FT /*tag= a

FT /product= "t1pe+"

PN US5871940-A.

XX 16-FEB-1999.

PF 13-JAN-1997; 97US-0782396.

PR 04-OCT-1994; 94US-0317880.

XX 13-JAN-1997; 97US-0782396.

XX (UYNV) UNIV NEW YORK STATE RES FOUND.

XX Feng G, Hall LM;

DR WPI: 1999-166632/14.

XX P-PSDB; AAM92459.

PT Screening for agents which modulate ion channel function - using

PT host cells transformed with nucleic acids encoding the Drosophila

XX melanogaster t1pe and para proteins

PS Disclosure; Fig 7A-B: 54pp; English.

XX This sequence encodes a Drosophila melanogaster t1pe+ protein which is

CC used in a method for screening for agents which modulate ion channel

CC function which uses host cells transformed with nucleic acid encoding

CC Drosophila melanogaster t1pe and para proteins. Co-expression of these

CC genes in the host cell, allows the formation of a functional voltage

CC dependent cation channel (VDCC) in the cell. The agents identified

CC can be used as pesticides for the control of Drosophila melanogaster or

CC other insects. They can also be used to screen pharmaceutical agents for

CC their neuroprotective affect against e.g. hypoxia, ischaemia, stroke and

XX head trauma.

XX Sequence 3954 BP; 1293 A; 960 C; 867 G; 834 T; 0 other;

Query Match 64.0%; Score 16; DB 20; Length 3954;

Best Local Similarity 79.2%; Pred. NO. 1.9e+02;

Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cattgatcctcgtcaggtgcag 25

||||| ||||| ||||| |||||

DB 163 CCTTCTATCCTTGCACGTCGAG 140

RESULT 32

AA094735/c

ID AA094735 standard; DNA: 4928 BP.

XX AA094735;

DT 02-APR-1996 (first entry)

DE Lycopene cyclase gene.

KW Lycopene cyclase; herbicide resistant plant; Synechococcus;

KW Synechocystis; ss.

OS Synechocystis sp.

XX Key Location/Qualifiers

FT CDS 2029..3264

FT /*tag= a

FT /product= lycopene cyclase

PN JP07155189-A.

XX 20-JUN-1995.

PF 01-FEB-1994; 94JP-0028793.

PR 25-OCT-1993; 93US-0142195.

XX (UYMA-) UNIV MARYLAND BALTIMORE.

DR WPI: 1995-250738/33.

XX P-PSDB; AAR76980.

PT Novel lycopene cyclase gene - useful for production of herbicide

PT resistant plants.

PS Claim 1; Page 12-15; 20pp; Japanese.

CC The DNA encodes a lycopene cyclase. The DNA is derived from Synechococcus

CC or Synechocystis, and is useful for the production of herbicide resistant

CC plants.

XX Sequence 4928 BP; 1046 A; 1415 C; 1394 G; 1073 T; 0 other;

QY 2 cattgatcctcgtcaggtgcag 25

||||| ||||| ||||| |||||

DB 4907 CAGCTCCATCATCTGTCAGGTCGAG 4884

RESULT 33

AAAT40205/c

ID AAAT40205 standard; DNA: 4928 BP.

XX AAAT40205;

DT 17-DEC-1996 (first entry)

DE Cyanobacterial lycopene cyclase lcy gene.

KW Lycopene cyclase; carotenoid; beta-carotene; herbicide resistance;

KW transgenic plant; crop protection; pigmentation; lcy gene; ss.

OS Synechococcus sp. PCC7942 (Anacystis nidulans R2).

XX Key Location/Qualifiers

FT CDS 2029..3264

FT /*tag= a

PN WO9628014-A1.

PD 19-SEP-1996.
 XX
 XX 05-MAR-1996; 96WO-US03044.
 XX
 XX 07-MAR-1995; 95US-0399561.
 XX
 PA (YISS) YISSUM RES & DEV CO.
 XX
 PI Cunningham FX, Gantt E, Hirschberg J;
 XX
 DR WPI: 1996-433414/43.
 DR P-PSDB; AAW00276.
 XX
 PT DNA encoding lycopen cyclase or mutants resistant to herbicides
 PT used to control carotenoid biosynthesis and to produce herbicide
 XX resistant plants and algae
 XX
 PS Claim 2; Page 33-37; 59pp; English.
 XX
 CC The lcy gene (AA140205) of *Synechococcus* sp. PCC 7942 codes for
 CC lycopene synthase (AAW00276), an enzyme that catalyses the prodn. of
 CC beta-carotene from lycopene. It was isolated from a genomic
 CC library of MPYA-resistant *Synechococcus* mutant Mr-5, by testing
 CC wild-type transformants for MPYA resistance, and analysis of the
 CC insert DNA of an isolated clone. The gene was used to isolate
 CC lycopene synthase cDNAs from tobacco and tomato (AA140206-07). The
 CC genes can be used to control carotenoid biosynthesis in bacteria,
 CC algae and plants e.g. to improve nutritional value (e.g. beta-
 CC carotene prodn.); pharmacological properties or appearance.
 CC Antisense sequences inhibit synthesis of the enzyme, resulting
 CC in increased red pigmentation. Mutant lcy genes can be used to
 CC impart herbicide resistance to transgenic plants.
 CC
 SQ Sequence 4928 BP; 1047 A; 1411 C; 1396 G; 1074 T; 0 other;

Query Match 64.0%; Score 16; DB 17; Length 4928;
 Best Local Similarity 79.2%; Pred. No. 1.9e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 cattcgatcctgcagtgagcag 25
 || ||||| ||||| ||||| |||||
 DB 4907 CAGCTCCATCATCTCAGGTTTCAG 4884

RESULT 34
 AAX13243/C
 ID AAX13243 standard; DNA; 11854 BP.
 XX
 AC AAX13243;
 XX
 DT 19-MAR-1999 (first entry)
 XX
 DE Enterococcus faecalis genome config SEQ ID NO:306.
 XX
 KM Enterococcus faecalis; config; detection; Enterococcal infection;
 KM vaccine; attenuation; computer readable medium; ds.
 OS Enterococcus faecalis.
 XX
 OS WO9850555-A2.
 PN
 XX
 PD 12-NOV-1998.
 XX
 PF 04-MAY-1998; 98WO-US08985.
 XX
 PR 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Dillon PJ, Kunsch CA;

XX
 DR WPI: 1999-045171/04.
 XX
 XX New isolated Enterococcus faecalis polynucleotides and polypeptides
 PT - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT infection.
 XX
 PS Claim 1; Page 1370-1376; 2084pp; English.
 XX
 CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC AAX12938 to AAX13919 represent these nucleotide sequences which are
 CC primary nucleotide sequences, also known as configs. The computer-based
 CC system can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.
 CC
 SQ Sequence 11854 BP; 4319 A; 1643 C; 2461 G; 3424 T; 7 other;

Query Match 64.0%; Score 16; DB 20; Length 11854;
 Best Local Similarity 79.2%; Pred. No. 2.2e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 cattcgatcctgcagtgagcag 25
 ||||| ||||| ||||| ||||| |||||
 DB 7291 CATTCCATCTTCATCAGTTACAG 7268

RESULT 35
 AAX20599/C
 ID AAX20599 standard; DNA; 20757 BP.
 XX
 AC AAX20599;
 XX
 DT 05-MAY-1999 (first entry)
 XX
 DE Polynucleotide sequence from the genome of *Treponema pallidum*.
 XX
 KM *Treponema pallidum* infection; syphilis; *Borrelia* infection; animal;
 KM enzyme production; ds.
 XX
 OS *Treponema pallidum*.
 XX
 PN WO9859034-A2.
 XX
 PD 30-DEC-1998.
 XX
 PF 23-JUN-1998; 98WO-US13041.
 XX
 PR 24-JUN-1997; 97US-0050667.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fraser CM;
 XX
 DR WPI: 1999-081273/07.
 XX
 PT New isolated *Treponema pallidum* nucleic acids - used to develop
 PT products for the detection, diagnosis, characterisation, prevention
 PT and therapy of T. pallidum infections, particularly syphilis
 XX
 PS Claim 1; Page 661-672; 1150pp; English.
 XX
 CC AAX20500-21243 represent polynucleotide sequences from the genome of

CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterization, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
XX
SQ Sequence 20757 BP; 4912 A; 6068 C; 5301 G; 4457 T; 19 other;
Query Match 64.0%; Score 16; DB 20; Length 20757;
Best Local Similarity 79.2%; Pred. No. 2.4e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 cattcgactcctgcagtgacg 25
|| ||||| || ||||| ||
Db 5481 CAGCTCGATCGCGAAGGTGCAC 5458
RESULT 36
AAF21105
ID AAF21105 standard; DNA: 143068 BP.
XX
AC AAF21105;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2672.
XX
KM Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KM human; airway disorder; bronchoconstriction; lung inflammation;
KM surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KM immunosuppressive; antiallergic; hypotensive; cytostatic;
KM surfactant hypoproduction; pulmonary obstruction; impeded respiration;
KM respiratory distress syndrome; pulmonary vasoconstriction; asthma; RDS;
KM pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KM chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KM cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
OS WPI: 2000-679539/66.
XX
DR Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure: Page 924-957; 1592pp; English.
XX
CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (1) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (1) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiallergic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (1) can be used to down-regulate the
CC expression and/or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,

CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
SQ Sequence 143068 BP; 41194 A; 30122 C; 32403 G; 39349 T; 0 other;
Query Match 64.0%; Score 16; DB 21; Length 143068;
Best Local Similarity 79.2%; Pred. No. 2.9e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 cattcgactcctgcagtgacg 25
|| ||||| || || ||||| ||
Db 86450 cattcgactcctgcagtgacg 86473
RESULT 37
AAF21272
ID AAF21272 standard; DNA: 143068 BP.
XX
AC AAF21272;
XX
DT 14-MAR-2001 (first entry)
XX
DE Human low adenosine antisense oligonucleotide related sequence #2839.
XX
KM Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KM human; airway disorder; bronchoconstriction; lung inflammation;
KM surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KM immunosuppressive; antiallergic; hypotensive; cytostatic;
KM surfactant hypoproduction; pulmonary obstruction; impeded respiration;
KM respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KM pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KM chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KM cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200062736-A2.
XX
PD 26-OCT-2000.
XX
PF 24-MAR-2000; 2000WO-US08020.
XX
PR 06-APR-1999; 99US-0127958.
XX
PA (UYEC-) UNIV EAST CAROLINA.
PA (NYCE/) NYCE J W.
XX
PI Nyce JW;
XX
OS WPI: 2000-679539/66.
XX
DR Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX

PS Disclosure: Page 1186-1219; 1592pp; English.

CC The present invention describes low adenosine (A) content antisense
 CC oligonucleotides and compositions (I) comprising them. In the antisense
 CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
 CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
 CC immunosuppressive, antiasthmatic, hypotensive and cytosolic activities.
 CC The antisense oligonucleotides and (I) can be used to down-regulate the
 CC expression and/or activity of target polypeptides associated with
 CC lung/respiratory disorders and malignancies, such as stimulating and
 CC activating peptide factors and transmitters, transcription factors,
 CC immunoglobulins and antibodies, antibody receptors, cytokines and
 CC chemokines, endogenously produced specific and non-specific enzymes,
 CC binding proteins, adhesion molecules and their receptors, cytokine and
 CC chemokine receptors, adenosine receptors, bradykinin receptors, central
 CC nervous system (CNS) and peripheral nervous and non-nervous system
 CC receptors, CNS and peripheral nervous and non-nervous system peptide
 CC transmitters, defensins, growth factors, vasoactive peptides and
 CC receptors, binding proteins and malignancy associated proteins. The
 CC antisense oligonucleotides may be used in this way to treat disorders
 CC including respiratory obstruction (especially pulmonary obstruction
 CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
 CC and/or surfactant hypoproduction which are associated with a disease or
 CC condition selected from pulmonary vasoconstriction, inflammation,
 CC allergies, asthma, impaired respiration, respiratory distress syndrome
 CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
 CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
 CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
 CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
 CC fragments and antisense oligonucleotides used in the exemplification of
 CC the present invention.

CC Sequence 143068 BP: 41194 A; 30122 C; 32403 G; 39349 T; 0 other;

DB 86450 cattcgatcctcgtagtgtag 25
 ||||| || |||||
 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match 64.0%; Score 16; DB 21; Length 143068;
 Best Local Similarity 79.2%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

DB 86450 cattcgatcctcgtagtgtag 25
 ||||| || |||||
 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

RESULT 38
 ID AAA34983 standard; DNA; 143068 BP.
 AC AAA34983;
 DT 28-JUL-2000 (first entry)

DE Human adenosine receptor related polynucleotide SEQ ID NO:2672.
 DE
 XX
 XX
 KW Human: adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytosolic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.
 OS
 PN WO200009525-A2.
 XX
 XX 24-FEB-2000.
 PD
 XX
 XX 03-AUG-1999; 99WO-US17712.
 PF
 XX 03-AUG-1998; 98US-0095212.
 PR
 XX

PA (UYEC-) UNIV EAST CAROLINA.
 XX
 XX Nye JW;
 PI
 XX
 DR WPI: 2000-205971/18.
 XX
 PT New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers

PS Disclosure: Page 851-882; 1343pp; English.

CC The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
 CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiasthmatic, cytosolic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC the ONS reduces side effects. The A-containing ONS break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing the
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONS from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

CC Sequence 143068 BP: 41194 A; 30122 C; 32402 G; 39350 T; 0 other;

DB 86450 cattcgatcctcgtagtgtag 25
 ||||| || |||||
 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Query Match 64.0%; Score 16; DB 21; Length 143068;
 Best Local Similarity 79.2%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

DB 86450 cattcgatcctcgtagtgtag 25
 ||||| || |||||
 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

RESULT 39
 ID AAA35150 standard; DNA; 143068 BP.
 AC AAA35150;
 DT 28-JUL-2000 (first entry)

DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:24.
 DE
 XX
 XX
 KW Human: adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytosolic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

OS Homo sapiens.
 OS
 PN WO200009525-A2.


```

FT      /product= "protein of unknown function"
FT      /note= "homologous to the FixX gene"
FT      /tag= l
FT      /standard_name= "ORF K12"
FT      /product= "protein of unknown function"
FT      /tag= m
FT      /standard_name= "ORF K13"
FT      /product= "ferrodoxin/ferrodoxin-like protein"
FT      /note= "homologous to the FdxN gene"
FT      /tag= n
FT      /standard_name= "ORF K14"
FT      /gene= "nifB"
FT      /product= "protein involved in Fmo co-factor
FT      biosynthesis"
FT      /tag= o
FT      /standard_name= "ORF K15"
FT      /gene= "nifA"
FT      /product= "positive regulator of nif, fix and other
FT      genes"
FT      /tag= p
FT      /standard_name= "ORF K16"
FT      /gene= "fixX"
FT      /product= "protein required for nitrogenase activity"
FT      /tag= q
FT      /standard_name= "ORF K17"
FT      /gene= "fixC"
FT      /product= "protein required for nitrogenase activity"
FT      /tag= r
FT      /standard_name= "ORF K18"
FT      /gene= "fixB"
FT      /product= "protein required for nitrogenase activity"
FT      /tag= s
FT      /standard_name= "ORF K19"
FT      /gene= "fixA"
FT      /product= "protein required for nitrogenase activity"
FT      /tag= t
FT      /standard_name= "ORF K20"
FT      /product= "protein of unknown function"
FT      /tag= u
FT      /standard_name= "ORF K21"
FT      /product= "protein of unknown function"
FT      /tag= v
FT      /standard_name= "ORF K22"
FT      /product= "ferrodoxin-like protein"
FT      /note= "homologous to the NifQ gene"
FT      /tag= w
FT      /standard_name= "ORF K23"
FT      /gene= "dctA"
FT      /product= "C4-dicarboxylate transport protein"
FT      /note= "homologous to the DctA gene"
FT      /tag= x
FT      /standard_name= "ORF L1"
FT      /product= "cytochrome P450-like protein"
FT      /note= "homologous to the CamC gene"
FT      /tag= y
FT      /standard_name= "ORF L2"
FT      /product= "gamma-hexachlorocyclohexane-dechlorinase-like
FT      protein"
FT      /note= "homologous to the LinA gene"

```

```

FT      CDS      448497..450203
FT      /tag= z
FT      /standard_name= "ORF L3"
FT      /product= "putative protein with degradative function"
FT      /tag= aa
FT      /standard_name= "ORF L4"
FT      /product= "luciferase alpha-subunit-like protein"
FT      /note= "homologous to the LuxA gene"
FT      /tag= ab
FT      /standard_name= "ORF L6"
FT      /gene= "nifD"
FT      /product= "alpha-subunit of Fmo protein of nitrogenase"
FT      /tag= ac
FT      /standard_name= "ORF L7"
FT      /gene= "nifK"
FT      /product= "beta-subunit of Fmo protein of nitrogenase"
FT      /tag= ad
FT      /standard_name= "ORF L8"
FT      /product= "protein involved in Fmo co-factor
FT      biosynthesis"
FT      /tag= ae
FT      /standard_name= "ORF L9"
FT      /product= "protein involved in Fmo co-factor
FT      biosynthesis"
FT      /tag= af
FT      /standard_name= "ORF L10"
FT      /product= "protein of unknown function"
FT      /note= "homologous to the NifX gene"
FT      /tag= ag
FT      /standard_name= "ORF L11"
FT      /product= "protein of unknown function"
FT      /tag= ah
FT      /standard_name= "ORF L12"
FT      /product= "protein similar to part of the Fe protein
FT      of nitrogenase"
FT      /note= "homologous to the NifH gene"
FT      /tag= ai
FT      /standard_name= "ORF L13"
FT      /product= "protein of unknown function"
FT      /tag= aj
FT      /standard_name= "ORF L14"
FT      /product= "peptidase-like protein"
FT      /note= "homologous to the bi-MPP gene"
FT      /tag= ak
FT      /standard_name= "ORF L15"
FT      /product= "processing protease-like protein"
FT      /note= "homologous to the PP gene"
FT      CDS      46590..467021

```

```

Query Match      64.0%; Score 16; DB 19; Length 534720;
Best Local Similarity 79.2%; Pred. No. 3, le+02;
Matches 19; Conservativity 0; Mismatches 5; Indels 0; Gaps 0;
QY      1 gcatttcgactcctcgtcaggtgca 24
      ||||| ||||| ||||| |||||
Db 461054 GCATTTCGACCTCTGCCCTGTGGA 461031

```

RESULT 43
AAV30459/c

```

ID  AAV30459 standard; DNA; 536165 BP.
XX
AC  AAV30459;
XX
DT  06-JUL-1999 (first entry)
XX
DE  Rhizobium species symbiotic plasmid pNGR234.
XX
KM  Symiosis: open reading frame; ORF: plasmid; vector: transportation;
KM  degradation; metabolism; host range; nitrogen fixation; nodulation;
KM  legume; plant; ds.
XX
OS  Rhizobium sp.
XX
PN  WO9802560-A2.
XX
PD  22-JAN-1998.
XX
PF  10-JUL-1997; 97WO-1B00950.
XX
PR  20-MAY-1997; 97GB-0010395.
PR  12-JUL-1996; 96EP-0730001.
XX
PA  (MOLE-) INST MOLECULAR BIOTECHNOLOGY.
PA  (BIOL-) LAB BIOLOGIE MOLECULAIRE PLANTES SUPERIE.
XX
PI  Broughton WJ, Freiberg CB, Perret XP, Rosenthal A;
XX  WPI; 1998-110606/10.
XX
DR  New isolated symbiotic plasmid from Rhizobium sp. NGR234 - used to
PT  develop products for modifying plant characteristics, e.g. nitrogen
PT  fixation, synthesis of compounds and stress response
XX
PS  Claim 1; Fig 3; 228bp; English.
XX
CC  This is the nucleotide sequence of the plasmid pNGR234a isolated from
CC  Rhizobium sp. NGR234. Open reading frames (ORF) derivable from the
CC  nucleotide sequence are claimed. The nucleotide sequences or ORFs can
CC  be used e.g. in the transportation of compounds to and from an organism
CC  which is a host to at least one of the nucleotide sequences, ORFs or
CC  proteins, the degradation and/or metabolism of organic, inorganic,
CC  natural or xenobiotic substances in a host organism or the modification
CC  of the host range, nitrogen fixation abilities; for obtaining a synthetic
CC  minimal set of ORFs required for functional Rhizobium-legume symbiosis,
CC  especially for nodulation efficiency on host plants.
XX
SQ  Sequence 536165 BP; 111291 A; 155755 C; 157864 G; 111255 T; 0 other;

Query Match          64.0%; Score 16; DB 19; Length 536165;
Best Local Similarity 79.2%; Pred. No. 3.1e+02;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy  1 gcaattcgatcctcgtcaggtcga 24
    ||||| ||||| |||||
Db  461054 GCATTTCGAGCCTCGCCCTGTGCA 461031

RESULT 44
AAF28005
ID  AAF28005 standard; cDNA; 220 BP.
XX
AC  AAF28005;
XX
DT  08-MAY-2001 (first entry)
XX
DE  Human TR13 coding sequence fragment SEQ ID NO: 14.
XX
KM  Human: tumour necrosis factor receptor; TR13; TR14; infection;
KM  cancer; autoimmune disease; allergy; inflammatory disease;
KM  graft rejection; apoptosis; cardiovascular disease; aneurysm; ss.
XX

```

```

OS  Homo sapiens.
XX
PN  WO200105834-A1.
XX
PD  25-JAN-2001.
XX
PF  14-JUL-2000; 2000WO-US19343.
XX
PR  16-JUL-1999; 99US-0144087.
PR  18-AUG-1999; 99US-0149450.
PR  20-AUG-1999; 99US-0149712.
PR  10-SEP-1999; 99US-0153089.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Ruben SM, Ni J, Young PE;
XX
DR  WPI; 2001-112682/12.
XX
PT  Nucleic acids encoding 2 human tumor necrosis factor receptor
PT  polypeptides ((TR13) and (TR14)), useful for the prevention, diagnosis
PT  and treatment of, e.g. cancers, acquired immune deficiency syndrome and
PT  hypohidrotic ectodermal dysplasia -
XX
PS  Claim 31; Page 382; 418bp; English.
XX
CC  The present invention provides the protein and coding sequences of the
CC  human tumour necrosis factor receptors TR13 and TR14. These sequences are
CC  useful in the diagnosis and treatment of many diseases, including cancer,
CC  autoimmune diseases, cardiovascular disorders, allergies,
CC  neurodegenerative diseases, graft rejection, inflammation, aneurysms and
CC  infections.
XX
SQ  Sequence 220 BP; 43 A; 63 C; 57 G; 56 T; 1 other;

```

```

Query Match          63.2%; Score 15.8; DB 22; Length 220;
Best Local Similarity 89.5%; Pred. No. 1.6e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy  7 cgatcctcgtcaggtcag 25
    ||||| ||||| |||||
Db  59 ccacccggtcaggtcag 77

RESULT 45
AAF28044
ID  AAF28044 standard; cDNA; 220 BP.
XX
AC  AAF28044;
XX
DT  08-MAY-2001 (first entry)
XX
DE  Human TR13 coding sequence fragment SEQ ID NO: 55.
XX
KM  Human: tumour necrosis factor receptor; TR13; TR14; infection;
KM  cancer; autoimmune disease; allergy; inflammatory disease;
KM  graft rejection; apoptosis; cardiovascular disease; aneurysm; ss.
XX
OS  Homo sapiens.
XX
PN  WO200105834-A1.
XX
PD  25-JAN-2001.
XX
PF  14-JUL-2000; 2000WO-US19343.
XX
PR  16-JUL-1999; 99US-0144087.
PR  18-AUG-1999; 99US-0149450.
PR  20-AUG-1999; 99US-0149712.
PR  10-SEP-1999; 99US-0153089.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX

```


Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred No. 0.0013;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatttcgactcctgcagtgacg 25
|||||
DB 99 GCATTTCGATCCTCCTCAGTGACG 123

RESULT 2

US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5869719artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846.338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatttcgactcctgcagtgacg 25
|||||
DB 99 GCATTTCGATCCTCCTCAGTGACG 123

RESULT 3

US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411.768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: pRO30A-15/9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 117..1157
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start=117
; OTHER INFORMATION: /product= "Biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioB"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2295..3050
; OTHER INFORMATION: /codon_start= 2295
; OTHER INFORMATION: /function= "Involved in pimeloyl-CoA synthesis"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /gene= "bioc"
; OTHER INFORMATION: /number= 3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3750..5039
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 3750
; OTHER INFORMATION: /EC_number= 2.6.1.62
; OTHER INFORMATION: /product= "DAPA synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioA"
; OTHER INFORMATION: /number= 5
; OTHER INFORMATION: /standard_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
; OTHER INFORMATION: /aminoacids="

LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter ptac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter ptac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "biob RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter ptac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1
Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gatttcgactctgctcgggagcag 25
DB 215 GCATTTCGATCCTCGTCAGGTCAG 239
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pB030A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biof"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biof"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "biof RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biob RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1

FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgcacccctgcagtgacg 25
|||||
Db 215 GCATTCGATCCTCGTCAGGTGACG 239

RESULT 5
US-08-726-725-5
Sequence 5, Application US/08726725
Patent No. 5773290
GENERAL INFORMATION:
APPLICANT: Gould, Michael N.
TITLE OF INVENTION: MAMMARY GLAND-SPECIFIC PROMOTERS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726.725
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Baker, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 960296.93863
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5709
TELEFAX: (414) 271-3552
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 876 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-726-725-5

Query Match 68.8%; Score 17.2; DB 1; Length 876;
Best Local Similarity 86.4%; Pred. No. 9.5;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 cattcgatcctgcagtgacg 23
|||||
Db 303 CACTTCATCCTCGTCAGGTGACG 324

RESULT 6
US-08-317-880-3/C
Sequence 3, Application US/08317880
Patent No. 5593862
GENERAL INFORMATION:
APPLICANT: Hall, Linda M.
TITLE OF INVENTION: Proteins for Cation Channel Expression
FILING DATE: 04-OCT-1994
ATTORNEY/AGENT INFORMATION:

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,880
FILING DATE: 04-OCT-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Rogalsky, Peter
REGISTRATION NUMBER: 38,601
REFERENCE/DOCKET NUMBER: 19226/610 (R-5261)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1634
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1045 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-317-880-3

Query Match 64.0%; Score 16; DB 1; Length 1045;
Best Local Similarity 79.2%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cattcgatcctgcagtgacg 25
|||||
Db 163 CATTTCATCCTCGTCAGGTGACG 140

RESULT 7
US-08-782-396-3/C
Sequence 3, Application US/08782396
Patent No. 5871940
GENERAL INFORMATION:
APPLICANT: Hall, Linda M.
TITLE OF INVENTION: PROTEINS FOR CATION CHANNEL EXPRESSION
FILING DATE: 13-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/782,396
FILING DATE: 13-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/782,396
FILING DATE: 13-JAN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

```

? NAME: Rogalskyj, Peter
? REGISTRATION NUMBER: 38,601
? REFERENCE/DOCKET NUMBER: 199226/611 (R-5261B)
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (7716) 263-1634
? TELEFAX: (7716) 263-1600
? INFORMATION FOR SEQ ID NO: 3:
?
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1045 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? OS-08-782-396-3

```

Query Match	64.0%;	Score 16;	DB 2;	Length 1045;
Best Local Similarity	79.2%;	Pred. No. 39;		
Matches 19; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0

```

Oy      2 cattcgatccctcgtagtgag 25
          | ||| | ||| | | ||| |||
Db     163 CCTTCTATCCTGCCACGTGCAG 140

```

```

1 RESULT 8
2 US-09-082-088-1/c
3 : Sequence 1, Application US/09082088
4 : Patent No. 6130067
5 :
6 : GENERAL INFORMATION:
7 : APPLICANT: TSUI, PING
8 : TITLE OF INVENTION: HUMAN EDG3sb GENE
9 : NUMBER OF SEQUENCES: 2
10 : CORRESPONDENCE ADDRESS:
11 : ADDRESSEE: Rather & Prestia
12 : STREET: P.O. Box 980
13 : CITY: Valley Forge
14 : STATE: PA
15 : COUNTRY: USA
16 : ZIP: 19482
17 :
18 : COMPUTER READABLE FORM:
19 : MEDIUM TYPE: Diskette
20 : COMPUTER: IBM Compatible
21 : OPERATING SYSTEM: DOS
22 : SOFTWARE: FASTSEQ for Windows Version 2.0
23 : CURRENT APPLICATION DATA:
24 : APPLICATION NUMBER: US/09/082,088
25 : FILING DATE: 20-MAY-1998
26 : CLASSIFICATION:
27 : PRIOR APPLICATION DATA:
28 : APPLICATION NUMBER:
29 : FILING DATE:
30 : ATTORNEY/AGENT INFORMATION:
31 : NAME: Prestia, Paul F
32 : REGISTRATION NUMBER: 23,031
33 : REFERENCE/DOCKET NUMBER: GP-70453
34 : TELECOMMUNICATION INFORMATION:
35 : TELEPHONE: 610-407-0700
36 : TELEFAX: 610-407-0701
37 :
38 : TELEX:
39 :
40 : INFORMATION FOR SEQ ID NO: 1:
41 : SEQUENCE CHARACTERISTICS:
42 : LENGTH: 1137 base pairs
43 : TYPE: nucleic acid
44 : STRANDEDNESS: single
45 : TOPOLOGY: linear
46 :
47 : MOLECULE TYPE: CDNA
48 :
49 : US-09-082-088-1

```

Query Match	64.0%;	Score 16;	DB 3;	length 1137;
Best Local Similarity	79.2%;	Pred. NO. 40;		
Matches 19;	Conservative	0;	Mismatches 5;	Indels 0;
			Gaps	0

```
QY      2 cattcgatccctcgtcagltgcag 25
          ||||| ||||| ||||| |||||
Db     429 CAATTGATCATTTGTCAAGTGGCG 406
```

```

US-08-330-537-3/c
US-08-330-537-3
Sequence 3, Application US/08330537
Patent No. 5663048
GENERAL INFORMATION:
APPLICANT: WINKFEIN, ROBERT J.
APPLICANT: DIXON, GORDON H.
TITLE OF INVENTION: Y-CHROMOSOME SPECIFIC POLYNUCLEOTIDE
TITLE OF INVENTION: PROBES FOR PRENATAL SEXING
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: 699 PRINCE, P.O. BOX 1404
CITY: ALEXANDRIA
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/330.537
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/030,485
FILING DATE: 13-MAY-1993
APPLICATION NUMBER: PCT/CA91/00353
FILING DATE: 03-OCT-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CA 2,026,926
FILING DATE: 04-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: DILLAHUNTY, T. GENE
REGISTRATION NUMBER: 25,423
REFERENCE/DOCKET NUMBER: 028722-080
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-854-7400
TELEFAX: 415-854-8275
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3713 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-330-537-3

```

Query Match	64.08;	Score 16;	DB 1;	Length 3713;
Best Local Similarity	79.28;	Pred. No. 50;		
Matches 19;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;

QY 1 gcatttcgatacctcgtcagtgca 24
||| ||| ||| ||| ||| ||| |||
Db 1537 GCAGATCGGCTCTTGTCAAGAGCA 1514

RESULT 10
US-08-330-537-4/C
Sequence 4, Application US/08330537
Patent No. 5663048
GENERAL INFORMATION:
APPLICANT: WINKFEIN, ROBERT J.
APPLICANT: DIXON, GORDON H.
TITLE OF INVENTION: Y-CHROMOSOME SPECIFIC POLYNUCLEOTIDE

;; TITLE OF INVENTION: PROBES FOR PRENATAL SEXING
;; NUMBER OF SEQUENCES: 4
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
;; STREET: 699 PRINCE, P.O. BOX 1404
;; CITY: ALEXANDRIA
;; STATE: VIRGINIA
;; COUNTRY: U.S.A.
;; ZIP: 22313-1404
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/330,537
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 08/030,485
;; FILING DATE: 13-MAY-1993
;; APPLICATION NUMBER: PCT/CA91/00353
;; FILING DATE: 03-OCT-1991
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: CA 2,026,926
;; FILING DATE: 04-OCT-1990
;; ATTORNEY/AGENT INFORMATION:
;; NAME: DILLAHUNTY, T. GENE
;; REGISTRATION NUMBER: 25,423
;; REFERENCE/DOCKET NUMBER: 028722-080
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-854-7400
;; TELEFAX: 415-854-8275
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3713 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; US-08-330-537-4

Query Match 64.0%; Score 16; DB 1; Length 3713;
Best Local Similarity 79.2%; Pred. No. 50;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 gcattcgatctctgcagtgca 24
DB 1537 GCAGATCGGCTCTGTCAGGACGA 1514

RESULT 11
US-08-317-880-1/c
;; Sequence 1, Application US/08317880
;; Patent No. 5593862
;; GENERAL INFORMATION:
;; APPLICANT: Hall, Linda M.
;; TITLE OF INVENTION: Proteins for Cation Channel Expression
;; TITLE OF INVENTION: and Function
;; NUMBER OF SEQUENCES: 5
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
;; STREET: Clinton Square, P.O. Box 1051
;; CITY: Rochester
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 14603
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/317,880
;; FILING DATE: 04-OCT-1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Rogalsky, Peter
;; REGISTRATION NUMBER: 38,601
;; REFERENCE/DOCKET NUMBER: 19226/610 (R-5261)
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (716) 263-1600
;; TELEFAX: (716) 263-1634
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3954 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-317-880-1

Query Match 64.0%; Score 16; DB 1; Length 3954;
Best Local Similarity 79.2%; Pred. No. 51;
Matches 19; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cattcgatctctgcagtgca 25
DB 163 CCTTCTATCCTTGCACGTCAG 140

RESULT 12
US-08-782-396-1/c
;; Sequence 1, Application US/08782396
;; Patent No. 5871940
;; GENERAL INFORMATION:
;; APPLICANT: Hall, Linda M.
;; TITLE OF INVENTION: PROTEINS FOR CATION CHANNEL EXPRESSION
;; TITLE OF INVENTION: AND FUNCTION
;; NUMBER OF SEQUENCES: 5
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
;; STREET: Clinton Square, P.O. Box 1051
;; CITY: Rochester
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 14603
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; OPERATING SYSTEM: IBM PC compatible
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/782,396
;; FILING DATE: 13-JAN-1997
;; CLASSIFICATION: 435
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: US 08/317,880
;; FILING DATE: 04-OCT-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Rogalsky, Peter
;; REGISTRATION NUMBER: 38,601
;; REFERENCE/DOCKET NUMBER: 19226/611 (R-5261B)
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (716) 263-1634
;; TELEFAX: (716) 263-1600
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3954 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-782-396-1

TELEX:
? INFORMATION FOR SEQ ID NO: 71:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1139 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: LATRTUT02
? CLONE: 1404643
? US-08-933-750C-71

Query Match 62.4%; Score 15.6; DB 2; Length 1139;
Best Local Similarity 81.8%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 attgcagctcgcagtgca 24
|||||
Db 668 ATTGGATCCTCATCTGCTGTA 647

RESULT 16
US-09-234-613-71/c
? Sequence 71, Application US/09234613
? Patent No. 6132973
? GENERAL INFORMATION:
? APPLICANT: Lal, Preeti
? APPLICANT: Hillman, Jennifer L.
? APPLICANT: Bandman, Olga
? APPLICANT: Shah, Purvi
? APPLICANT: Au-Young, Janice
? APPLICANT: Yue, Henry
? APPLICANT: Guegler, Karl J.
? APPLICANT: Corley, Neil C.
? TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
? NUMBER OF SEQUENCES: 98
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Incyte Pharmaceuticals, Inc.
? STREET: 3174 Porter Drive
? CITY: Palo Alto
? STATE: CA
? COUNTRY: USA
? ZIP: 94304
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSeq for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/234,613
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US/08/933,750
? FILING DATE: September 23, 1997
? ATTORNEY/AGENT INFORMATION:
? NAME: Billings, Lucy J.
? REGISTRATION NUMBER: 36,749
? REFERENCE/DOCKET NUMBER: PF-0356 US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-855-0555
? TELEFAX: 415-845-4166
? TELEX:
? INFORMATION FOR SEQ ID NO: 71:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1139 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: LATRTUT02
? CLONE: 1404643

US-09-234-613-71

Query Match 62.4%; Score 15.6; DB 3; Length 1139;
Best Local Similarity 81.8%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 attgcagctcgcagtgca 24
|||||
Db 668 ATTGGATCCTCATCTGCTGTA 647

RESULT 17
US-07-841-646-8
? Sequence 8, Application US/07841646
? Patent No. 5266683
? GENERAL INFORMATION:
? APPLICANT: OPPERMAN, HERMANN
? APPLICANT: OKAYNAK, ENGIN
? APPLICANT: KUBERASAMPATH, THANGAVEL
? APPLICANT: RUEGER, DAVID C.
? APPLICANT: PANG, ROY H. L.
? TITLE OF INVENTION: OSTEOGENIC DEVICES
? NUMBER OF SEQUENCES: 33
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: TESTA, HURWITZ & THIBEAULT
? STREET: 53 STATE STREET
? CITY: BOSTON
? STATE: MASSACHUSETTS
? COUNTRY: U.S.A.
? ZIP: 02109
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/07/841,646
? FILING DATE: 19920221
? CLASSIFICATION: 530
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 810,560
? FILING DATE: 20-DEC-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 827,052
? FILING DATE: 28-JAN-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 660,162
? FILING DATE: 22-FEB-1991
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 621,988
? FILING DATE: 04-DEC-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 621,849
? FILING DATE: 04-DEC-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 616,374
? FILING DATE: 21-NOV-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 600,024
? FILING DATE: 18-OCT-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 599,543
? FILING DATE: 18-OCT-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 579,865
? FILING DATE: 07-SEP-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 569,920
? FILING DATE: 20-AUG-1990
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 483,913
? FILING DATE: 22-FEB-1990


```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..507
; OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
; OTHER INFORMATION: /product= "Opla"
; OTHER INFORMATION: /note= "Opla FUSION"
;
; US-07-841-646-8
;
Query Match 61.6%; Score 15.4; DB 1; Length 516;
Best Local Similarity 76.0%; Pred. No. 67;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
;
QY 1 gcatcgcacgcctcgcaggtcag 25
   | | | | | | | | | | | | | | |
Db 174 GGAATTCGATCTCTCACCAGAGCAG 198
;
RESULT 18
US-08-147-023-8
; Sequence 8, Application US/08147023
; Patent No. 5468645
; GENERAL INFORMATION:
; APPLICANT: OPPERMAN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H. L.
; TITLE OF INVENTION: OSTEOGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147.023
; FILING DATE: 21-FEB-1992
;

```

```

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 827,052
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 616,374
; FILING DATE: 21-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 600,024
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 599,543
; FILING DATE: 18-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 579,865
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 569,920
; FILING DATE: 20-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 483,913
; FILING DATE: 22-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 422,613
; FILING DATE: 17-OCT-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 315,342
; FILING DATE: 23-FEB-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 232,630
; FILING DATE: 15-AUG-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 179,460
; FILING DATE: 08-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-001CP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 516 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..507
; OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
; OTHER INFORMATION: /product= "Opla"
; OTHER INFORMATION: /note= "Opla FUSION"
;
US-08-147-023-8
;
Query Match 61.6%; Score 15.4; DB 1; Length 516;
Best Local Similarity 76.0%; Pred. No. 67;
;

```



```

1      APPLICATION NUMBER:  US 827,052
2      FILING DATE:  28-JAN-1992
3      PRIOR APPLICATION DATA:
4      APPLICATION NUMBER:  US 660,162
5      FILING DATE:  22-FEB-1991
6      PRIOR APPLICATION DATA:
7      APPLICATION NUMBER:  US 621,988
8      FILING DATE:  04-DEC-1990
9      PRIOR APPLICATION DATA:
10     APPLICATION NUMBER:  US 621,849
11     FILING DATE:  04-DEC-1990
12     PRIOR APPLICATION DATA:
13     APPLICATION NUMBER:  US 616,374
14     FILING DATE:  21-NOV-1990
15     PRIOR APPLICATION DATA:
16     APPLICATION NUMBER:  US 600,024
17     FILING DATE:  18-OCT-1990
18     PRIOR APPLICATION DATA:
19     APPLICATION NUMBER:  US 599,543
20     FILING DATE:  18-OCT-1990
21     PRIOR APPLICATION DATA:
22     APPLICATION NUMBER:  US 579,865
23     FILING DATE:  07-SEP-1990
24     PRIOR APPLICATION DATA:
25     APPLICATION NUMBER:  US 569,920
26     FILING DATE:  20-AUG-1990
27     PRIOR APPLICATION DATA:
28     APPLICATION NUMBER:  US 483,913
29     FILING DATE:  22-FEB-1990
30     PRIOR APPLICATION DATA:
31     APPLICATION NUMBER:  US 422,613
32     FILING DATE:  17-OCT-1989
33     PRIOR APPLICATION DATA:
34     APPLICATION NUMBER:  US 315,342
35     FILING DATE:  23-FEB-1989
36     PRIOR APPLICATION DATA:
37     APPLICATION NUMBER:  US 232,630
38     FILING DATE:  15-AUG-1988
39     PRIOR APPLICATION DATA:
40     APPLICATION NUMBER:  US 179,460
41     FILING DATE:  08-APR-1988
42     ATTORNEY/AGENT INFORMATION:
43     NAME:  PITCHER, EDMUND R.
44     REGISTRATION NUMBER:  27,829
45     REFERENCE/DOCKET NUMBER:  CRP-001CP6
46     TELECOMMUNICATION INFORMATION:
47     TELEPHONE:  617/248-7000
48     TELEFAX:  617/248-7100
49     INFORMATION FOR SEQ ID NO:  8:
50     SEQUENCE CHARACTERISTICS:
51     LENGTH:  516 base pairs
52     TYPE:  nucleic acid
53     STRANDEDNESS:  single
54     TOPOLOGY:  linear
55     MOLECULE TYPE:  cDNA
56     HYPOTHETICAL:  YES
57     ANTI-SENSE:  NO
58     FEATURE:
59     NAME/KEY:  CDS
60     LOCATION:  1..507
61     OTHER INFORMATION:  /function= "OSTEOGENIC PROTEIN"
62     OTHER INFORMATION:  /product= "Opia"
63     OTHER INFORMATION:  /note= "Opia FUSION"
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
```

Query	Match	Similarity	61.6%	Score 15.4	DB 2	Length 516
174	GGAATTGCATCTCTCACGAGGACAG	198	76.0%	Pred. No. 67		
	Matches 19	Conservative 0	Mismatches 6	Indels 0	Gaps 0	

```

RESULT 21
US-08-449-699A-8
: Sequence 8, Application US/08449699A
: Patent No. 5958441
: GENERAL INFORMATION:
: APPLICANT: OPPERMAN, HERMANN
: APPLICANT: OZKAYNAK, ENGIN
: APPLICANT: KUBERASAMPATH, THANCAVEL
: APPLICANT: RUEGER, DAVID C.
: APPLICANT: PANG, ROY H.L.
: TITLE OF INVENTION: ANTIBODIES TO OSTEOGENIC PROTEINS
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TESTA, HURWITZ & THIBEAULT
: STREET: 125 HIGH STREET
: CITY: BOSTON
: STATE: MASSACHUSETTS
: COUNTRY: U.S.A.
: ZIP: 02110
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/449,699A
: FILING DATE: 24-MAY-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/147,023
: FILING DATE: 01-NOV-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER, EDMUND R.
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: STK-001CP6CN
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/248-7000
: TELEFAX: 617/248-7100
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 516 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEetical: YES
: ANTI-SENSE: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..507
: OTHER INFORMATION: /function= "OSTEOGENIC PROTEIN"
: OTHER INFORMATION: /product= "Op1A"
: OTHER INFORMATION: /note= "Op1A fusion"
US-08-449-699A-8

Query Match 61.6%, Score 15.4; DB 2; Length 516;
Best Local Similarity 76.0%; Pred. NO. 67;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Oy 1 gcatcgcacctcgtcaggtgcag 25
| | ||||| ||| |||
Db 174 GGAATTCATCTCTCACGAGGCGAG 198

RESULT 22
US-09-193-377B-2
: Sequence 2, Application US/09193377B
: Patent No. 6221594
: GENERAL INFORMATION:
: APPLICANT: Burrell, Paul

```

```
; APPLICANT: Blackall, Linda
; APPLICANT: Keller, Jurg
; TITLE OF INVENTION: METHOD FOR THE DETECTION OF AQUATIC
; TITLE OF INVENTION: NITRITE OXIDISING MICROORGANISMS OF THE GENUS NITROSPIRA
; FILE REFERENCE: CULIN20.001AUS
; CURRENT APPLICATION NUMBER: US/09/193,377B
; CURRENT FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1407
; TYPE: DNA
; ORGANISM: Nitrospira
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1407)
; OTHER INFORMATION: n = A,T,C or G
US-09-193-377B-2
```

```
Query Match          61.6%; Score 15.4; DB 4; Length 1407;
Best Local Similarity 76.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 gcaattgcattcctgcagtgacg 25
    ||| | | ||||| ||||| |||
DB 976 gcaataccattcctgcagtgacg 1000
```

```
RESULT 23
US-09-193-377B-4
; Sequence 4, Application US/09193377B
; Patent No. 6221594
; GENERAL INFORMATION:
; APPLICANT: Burrell, Paul
; APPLICANT: Blackall, Linda
; APPLICANT: Keller, Jurg
; TITLE OF INVENTION: METHOD FOR THE DETECTION OF AQUATIC
; TITLE OF INVENTION: NITRITE OXIDISING MICROORGANISMS OF THE GENUS NITROSPIRA
; FILE REFERENCE: CULIN20.001AUS
; CURRENT APPLICATION NUMBER: US/09/193,377B
; CURRENT FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1420
; TYPE: DNA
; ORGANISM: Nitrospira
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1420)
; OTHER INFORMATION: n = A,T,C or G
US-09-193-377B-4
```

```
Query Match          61.6%; Score 15.4; DB 4; Length 1420;
Best Local Similarity 76.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 gcaattgcattcctgcagtgacg 25
    ||| | | ||||| ||||| |||
DB 958 gcaataccattcctgcagtgacg 982
```

```
RESULT 24
US-09-193-377B-1
; Sequence 1, Application US/09193377B
; Patent No. 6221594
; GENERAL INFORMATION:
; APPLICANT: Burrell, Paul
; APPLICANT: Blackall, Linda
; APPLICANT: Keller, Jurg
; TITLE OF INVENTION: METHOD FOR THE DETECTION OF AQUATIC
```

```
; TITLE OF INVENTION: NITRITE OXIDISING MICROORGANISMS OF THE GENUS NITROSPIRA
; FILE REFERENCE: CULIN20.001AUS
; CURRENT APPLICATION NUMBER: US/09/193,377B
; CURRENT FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1428
; TYPE: DNA
; ORGANISM: Nitrospira
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1428)
; OTHER INFORMATION: n = A,T,C or G
US-09-193-377B-1
```

```
Query Match          61.6%; Score 15.4; DB 4; Length 1428;
Best Local Similarity 76.0%; Pred. No. 82;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 1 gcaattgcattcctgcagtgacg 25
    ||| | | ||||| ||||| |||
DB 967 gcaataccattcctgcagtgacg 991
```

```
RESULT 25
US-08-199-776-1/C
; Sequence 1, Application US/08199776
; Patent No. 5594120
; GENERAL INFORMATION:
; APPLICANT: Brenner, Michael B.
; APPLICANT: Parker, Christina M.
; TITLE OF INVENTION: No. 5594120e1 Integrin alpha subunit
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/199,776
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3933 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL TYPE: Mucosal Lymphocyte
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 126..3662
```

FEATURE:
NAME/KEY: mat_peptide
LOCATION: 180..3659
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 126..179
US-08-199-776-1

Query Match 61.6%; Score 15.4; DB 1; Length 3933;
Best Local Similarity 76.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcaatcgatccctgcagtgagcag 25
|| ||| ||||| ||| |||||
Db 2778 GCCTTGCATCCTCTCACTGCAG 2754

RESULT 26
US-08-663-731-1/C
Sequence 1, Application US/08663731
Patent No. 6057423
GENERAL INFORMATION:
APPLICANT: Brenner, Michael B.
APPLICANT: Parker, Christina M.
TITLE OF INVENTION: No. 6057423e1 integrin alpha subunit
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/663,731
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/199,776
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7020
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3933 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: mucosal lymphocyte
FEATURE:
NAME/KEY: CDS
LOCATION: 126..3662
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 180..3659
FEATURE:
NAME/KEY: sig_peptide

LOCATION: 126..179
US-08-663-731-1

Query Match 61.6%; Score 15.4; DB 3; Length 3933;
Best Local Similarity 76.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcaatcgatccctgcagtgagcag 25
|| ||| ||||| ||| |||||
Db 2778 GCCTTGCATCCTCTCACTGCAG 2754

RESULT 27
US-08-879-338-1/C
Sequence 1, Application US/08879338A
Patent No. 6063906
GENERAL INFORMATION:
APPLICANT: Brenner, Michael B.
APPLICANT: Parker, Christina M.
TITLE OF INVENTION: Antibodies to No. 6063906e1 Integrin Alpha
FILE REFERENCE: B0801/7080/ERP
CURRENT APPLICATION NUMBER: US/08/879,338A
CURRENT FILING DATE: 1997-06-20
EARLIER APPLICATION NUMBER: US 08/663,731
EARLIER FILING DATE: 1996-06-14
EARLIER APPLICATION NUMBER: US 08/199,776
EARLIER FILING DATE: 1994-02-18
NUMBER OF SEQ ID NOS: 31
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 3933
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (126)...(3659)
NAME/KEY: mat_peptide
LOCATION: (180)...(3659)
NAME/KEY: sig_peptide
LOCATION: (126)...(179)
US-08-879-338-1

Query Match 61.6%; Score 15.4; DB 3; Length 3933;
Best Local Similarity 76.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 gcaatcgatccctgcagtgagcag 25
|| ||| ||||| ||| |||||
Db 2778 GCCTTGCATCCTCTCACTGCAG 2754

RESULT 28
PCT-US95-02044-1/C
Sequence 1, Application PC/TUS9502044
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Novel integrin alpha subunit
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield and Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02044
; FILING DATE: herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/199,776
; FILING DATE: 18 February 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3933 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL TYPE: mucosal lymphocyte
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 126..3662
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 180..3659
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 126..179
;
; PCT-US95-02044-1
;
Query Match 61.6%; Score 15.4; DB 5; Length 3933;
Best Local Similarity 76.0%; Pred. No. 1e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
;
QY 1 gcattcgatcctcgctgagtgca 25
; 111111111111111111111111
;
Db 2778 GCTTTGCATCCTCTCACTGCAG 2754
;
RESULT 29
US-09-051-019-1
; Sequence 1, Application US/09051019
; Patent No. 6103229
; GENERAL INFORMATION:
; APPLICANT: KAHMANN, Regine and QUADBECK-SEGER, Claudia
; TITLE OF INVENTION: Regulatory gene from Ustilago maydis
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Keil & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.2 Mb storage
; COMPUTER: IBM AT-compatible, Pentium processor
; OPERATING SYSTEM: Windows 98
; SOFTWARE: Wordperfect version 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/051,019
; FILING DATE: 31-MAR-1998
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8931 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-051-019-1
;
Query Match 61.6%; Score 15.4; DB 3; Length 8931;
Best Local Similarity 76.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
;
QY 1 gcattcgatcctcgctgagtgca 25
; 111111111111111111111111
;
Db 6146 GCCATTCTCTGCTGCTGAGGCGCAG 6170
;
RESULT 30
US-08-441-971-16/c
; Sequence 16, Application US/08441971
; Patent No. 6071693
; GENERAL INFORMATION:
; APPLICANT: Tai-An Cha
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,971
; FILING DATE: 16-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,653
; FILING DATE:
; APPLICATION NUMBER: US/07/881,528
; FILING DATE:
; APPLICATION NUMBER: 07/697,326
; FILING DATE: 8 May 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Janluk, Anthony J.
; REGISTRATION NUMBER: 29,809
; REFERENCE/DOCKET NUMBER: C0772/7000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 720-3500
; TELEFAX: (617) 720-2441
; TELEX: EZEKIEL
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ns5arg6
; US-08-441-971-16
;
Query Match 60.8%; Score 15.2; DB 3; Length 340;
Best Local Similarity 85.0%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
QY 1 gcattcgatcctcgctgagtgca 20
```

```

;
; LENGTH: 8931 base pairs
; TYPE: Nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-09-051-019-1
;
Query Match 61.6%; Score 15.4; DB 3; Length 8931;
Best Local Similarity 76.0%; Pred. No. 1.2e+02;
Matches 19; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
;
QY 1 gcattcgatcctcgctgagtgca 25
; 111111111111111111111111
;
Db 6146 GCCATTCTCTGCTGCTGAGGCGCAG 6170
;
RESULT 30
US-08-441-971-16/c
; Sequence 16, Application US/08441971
; Patent No. 6071693
; GENERAL INFORMATION:
; APPLICANT: Tai-An Cha
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,971
; FILING DATE: 16-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,653
; FILING DATE:
; APPLICATION NUMBER: US/07/881,528
; FILING DATE:
; APPLICATION NUMBER: 07/697,326
; FILING DATE: 8 May 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Janluk, Anthony J.
; REGISTRATION NUMBER: 29,809
; REFERENCE/DOCKET NUMBER: C0772/7000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 720-3500
; TELEFAX: (617) 720-2441
; TELEX: EZEKIEL
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ns5arg6
; US-08-441-971-16
;
Query Match 60.8%; Score 15.2; DB 3; Length 340;
Best Local Similarity 85.0%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
;
QY 1 gcattcgatcctcgctgagtgca 20
```

Db 81 GCAGTTCGAGCCTCCTCAGG 62

RESULT 31

US-08-221-653-16/c
; Sequence 16, Application US/08221653
; Patent No. 6190864

GENERAL INFORMATION:

APPLICANT: Tai-An Cha

TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 147

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wolf, Greenfield & Sacks, P.C.

STREET: 600 Atlantic Avenue

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch

OPERATING SYSTEM: MS-DOS Version 3.3

SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/221.653

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/881.528

FILING DATE:

APPLICATION NUMBER: 07/697.326

FILING DATE: 8 May 1991

ATTORNEY/AGENT INFORMATION:

NAME: Janluk, Anthony J.

REGISTRATION NUMBER: 29,809

REFERENCE/DOCKET NUMBER: C0772/7000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 720-3500

TELEFAX: (617) 720-2441

TELEX: EZEKIEL

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 340 nucleotides

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: ns5arg6

US-08-221-653-16

Query Match

Best Local Similarity 60.8%; Score 15.2; DB 4; Length 340;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcattgcattcctgcag 20

Db 81 GCAGTTCGAGCCTCCTCAGG 62

RESULT 32

US-08-442-144A-16/c

; Sequence 16, Application US/08442144A

; Patent No. 6214583

GENERAL INFORMATION:

APPLICANT: Tai-An Cha

APPLICANT: Eileen Beall

APPLICANT: Bruce Irvine

APPLICANT: Janice Kolberg

APPLICANT: Michael S. Urdea

TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 148

CORRESPONDENCE ADDRESS:

ADDRESSEE: Chilton Corporation

STREET: 4560 Horton Street

CITY: Emeryville

STATE: California

COUNTRY: USA

ZIP: 94608-2916

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 Inch

OPERATING SYSTEM: Windows NT

SOFTWARE: Microsoft Word 97

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/442.144A

FILING DATE: MAY 16, 1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/221.653

FILING DATE: APRIL 1, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Doreen Yalko Trujillo

REGISTRATION NUMBER: 35,719

REFERENCE/DOCKET NUMBER: CHIR-0121

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-568-3100

TELEFAX: 215-568-3439

TELEX:

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 340 Nucleotides

TYPE: Nucleic Acid

STRANDEDNESS: Single

TOPOLOGY: Linear

MOLECULE TYPE: DNA

ORIGINAL SOURCE:

INDIVIDUAL ISOLATE: ns5arg6

US-08-442-144A-16

Query Match

Best Local Similarity 60.8%; Score 15.2; DB 4; Length 340;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 gcattgcattcctgcag 20

Db 81 GCAGTTCGAGCCTCCTCAGG 62

RESULT 33

US-08-528-199-2/c

; Sequence 2, Application US/08528199

; Patent No. 5763228

GENERAL INFORMATION:

APPLICANT: KUBOTA, Michio

APPLICANT: TSUSAKI, Keiji

APPLICANT: SUGIMOTO, Toshiyuki

TITLE OF INVENTION: RECOMBINANT ENZYME FOR CONVERTING

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK

STREET: 419 Seventh Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 140:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 27 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE: OTHER INFORMATION: The letter "N" stands for the stem II
: OTHER INFORMATION: region of a HH ribozyme.
US-08-758-306-140

Query Match 60.0%; Score 15; DB 1; Length 27;
Best Local Similarity 75.0%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 cattgcacccctgcagtgagcag 25
||||| ||||| ||||| ||||| |||||
DB 26 CATTGTTTCATCAGTCAGTCAG 3

RESULT 36
US-08-758-306-716/C
: Sequence 716, Application US/08758306
: Patent No. 5807743
: GENERAL INFORMATION:
: APPLICANT: Slinchcomb, Dan T.
: APPLICANT: McSwiggen, James A.
: TITLE OF INVENTION: METHOD AND REAGENT FOR THE
: TITLE OF INVENTION: TREATMENT OF DISEASES
: TITLE OF INVENTION: ASSOCIATED WITH
: TITLE OF INVENTION: INTERLEUKIN-2 RECEPTOR
: NUMBER OF SEQUENCES: 1379
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071-2066
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: Fastseq Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/758,306
: FILING DATE: December 3, 1996
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 21/132
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 716:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 27 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE: OTHER INFORMATION: The letter "N" stands for the stem II
: OTHER INFORMATION: region of a HH ribozyme.

US-08-758-306-716

Query Match 60.0%; Score 15; DB 1; Length 27;
Best Local Similarity 75.0%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 cattgcacccctgcagtgagcag 25
||||| ||||| ||||| ||||| |||||
DB 26 CATTGTTTCATCAGTCAGTCAG 3

RESULT 37
US-08-758-306-1112/C
: Sequence 1112, Application US/08758306
: Patent No. 5807743
: GENERAL INFORMATION:
: APPLICANT: Slinchcomb, Dan T.
: APPLICANT: McSwiggen, James A.
: TITLE OF INVENTION: METHOD AND REAGENT FOR THE
: TITLE OF INVENTION: TREATMENT OF DISEASES
: TITLE OF INVENTION: ASSOCIATED WITH
: TITLE OF INVENTION: INTERLEUKIN-2 RECEPTOR
: NUMBER OF SEQUENCES: 1379
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071-2066
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: Fastseq Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/758,306
: FILING DATE: December 3, 1996
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 212/132
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 1112:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 27 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE: OTHER INFORMATION: The letter "N" stands for the stem II
: OTHER INFORMATION: region of a HH ribozyme.
US-08-758-306-1112

Query Match 60.0%; Score 15; DB 1; Length 27;
Best Local Similarity 75.0%; Pred. No. 58;
Matches 18; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 cattgcacccctgcagtgagcag 25
||||| ||||| ||||| ||||| |||||
DB 26 CATTGTTTCATCAGTCAGTCAG 3

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1667RP
US-08-998-416-1093

Query Match 60.0%; Score 15; DB 4; Length 707;
Best Local Similarity 78.3%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 catlccgacccgcagtcgca 24
||| |||| |||| |||| ||||
Db 167 CCTGAGATCGCGTCAGCTGCA 145

RESULT 41
US-08-557-128-3/C
Sequence 3, Application US/08557128
Patent No. 5849524
GENERAL INFORMATION:
APPLICANT: KONDO, Keiji
APPLICANT: KAJIWARA, Susumu
APPLICANT: MISAWA, No. 5849524ibiko
TITLE OF INVENTION: TRANSFORMATION SYSTEMS FOR THE YEAST
TITLE OF INVENTION: CANDIDA UTILIS AND THE EXPRESSION OF HETEROLOGOUS GENES
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,128
FILING DATE: 25-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/JP95/01005
FILING DATE: 25-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-129287
FILING DATE: 28-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-285823
FILING DATE: 26-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-135015
FILING DATE: 25-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 49441/108
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5500
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2330 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS

LOCATION: 1259..2059
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1259..2059
US-08-557-128-3

Query Match 60.0%; Score 15; DB 2; Length 2330;
Best Local Similarity 78.3%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 catlccgacccgcagtcgca 24
||| | |||| |||| |||| ||||
Db 121 CATTACTATCCACGTCGGGTGA 99

RESULT 42
US-07-938-782A-1
Sequence 1, Application US/07938782A
Patent No. 5525513
GENERAL INFORMATION:
APPLICANT: Chen, Jane J.
APPLICANT: London, Irving M.
TITLE OF INVENTION: DNA Encoding the Heme-Regulated
TITLE OF INVENTION: Eukaryotic Initiation Factor 2 alpha kinase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/938,782A
FILING DATE: 31-AUG-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: W0101CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)873-8794
TELEFAX: (404)873-8795
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2729 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Rabbit
CELL TYPE: Reticulocytes
FEATURE:
NAME/KEY: misc-feature
LOCATION: 113..2149
OTHER INFORMATION: /note="Expression of HRI
OTHER INFORMATION: mRNA in Human erythroid cells, using as the
OTHER INFORMATION: probe rabbit HRI cDNA from nucleotides 113
OTHER INFORMATION: to 2149."
FEATURE:
NAME/KEY: misc-feature
LOCATION: 229..249

```
OTHER INFORMATION: /note= "Primer used in the
OTHER INFORMATION: amplification of human HRI cDNA sequence
OTHER INFORMATION: using the rabbit HRI cDNA sequence."
FEATURE:
NAME/KEY: misc feature
LOCATION: 543..560
OTHER INFORMATION: /note= "Primer used in the
OTHER INFORMATION: amplification of human HRI cDNA sequence
OTHER INFORMATION: using the rabbit HRI cDNA sequence."
FEATURE:
NAME/KEY: misc feature
LOCATION: 448..468
OTHER INFORMATION: /note= "Primer used in the
OTHER INFORMATION: amplification of human HRI cDNA sequence
OTHER INFORMATION: using the rabbit HRI cDNA sequence."
FEATURE:
NAME/KEY: misc feature
LOCATION: 1009..1031
OTHER INFORMATION: /note= "Primer used in the
OTHER INFORMATION: amplification of a human HRI cDNA sequence
OTHER INFORMATION: using the rabbit HRI cDNA sequence."
PUBLICATION INFORMATION:
AUTHORS: Chen, Jane J.
TITLE: Cloning of the cDNA of the heme-
TITLE: regulated eukaryotic initiation factor
TITLE: 2alpha (eIF-2alpha)kinase of rabbit reticulocytes:
TITLE: Homology to yeast GCN2 protein kinase and human
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 88
PAGES: 7729-7733
DATE: September-1991
US-07-938-782A-1

Query Match 60.0%; Score 15; DB 1; Length 2729;
Best Local Similarity 78.3%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 attcgatctctgcaggtgcag 25
Db 2059 ATTTCAGCCTCCTGAGCTTCAG 2081

RESULT 43
US-08-630-524-1
Sequence 1, Application US/08630524
Patent No. 5690930
GENERAL INFORMATION:
APPLICANT: Chen, Jane J.
APPLICANT: London, Irving M.
TITLE OF INVENTION: DNA Encoding the Heme-Regulated
TITLE OF INVENTION: Eukaryotic Initiation Factor 2 alpha kinase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,524
FILING DATE: 10-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
```

```
APPLICATION NUMBER: US 07/938,782
FILING DATE: 31-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: W0101CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)873-8794
TELEFAX: (404)873-8795
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2729 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Rabbit
CELL TYPE: Reticulocytes
FEATURE:
NAME/KEY: misc feature
LOCATION: 113..2149
OTHER INFORMATION: /note= "Expression of HRI
OTHER INFORMATION: mRNA in Human erythroid cells, using as the
OTHER INFORMATION: probe rabbit HRI cDNA from nucleotides 113
OTHER INFORMATION: to 2149."
FEATURE:
NAME/KEY: misc feature
LOCATION: 229..249
OTHER INFORMATION: /note= "Primer used in the
OTHER INFORMATION: amplification of human HRI cDNA sequence
OTHER INFORMATION: using the rabbit HRI cDNA sequence."
FEATURE:
NAME/KEY: misc feature
LOCATION: 543..560
OTHER INFORMATION: /note= "Primer used in the
OTHER INFORMATION: amplification of human HRI cDNA sequence
OTHER INFORMATION: using the rabbit HRI cDNA sequence."
FEATURE:
NAME/KEY: misc feature
LOCATION: 448..468
OTHER INFORMATION: /note= "Primer used in the
OTHER INFORMATION: amplification of human HRI cDNA sequence
OTHER INFORMATION: using the rabbit HRI cDNA sequence."
FEATURE:
NAME/KEY: misc feature
LOCATION: 1009..1031
OTHER INFORMATION: /note= "Primer used in the
OTHER INFORMATION: amplification of a human HRI cDNA sequence
OTHER INFORMATION: using the rabbit HRI cDNA sequence."
PUBLICATION INFORMATION:
AUTHORS: Chen, Jane J.
AUTHORS: London, Irving M.
TITLE: Cloning of the cDNA of the heme-
TITLE: regulated eukaryotic initiation factor
TITLE: 2alpha (eIF-2alpha)kinase of rabbit reticulocytes:
TITLE: Homology to yeast GCN2 protein kinase and human
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 88
PAGES: 7729-7733
DATE: September-1991
US-08-630-524-1

Query Match 60.0%; Score 15; DB 1; Length 2729;
Best Local Similarity 78.3%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 3 attcgatctctgcaggtgcag 25
```

Db 2059 ATTCAAGCCTCCTGAGGTTGAG 2081

RESULT 44

PCT-US93-08131-1
Sequence 1, Application PC/TUS9308131
GENERAL INFORMATION:
APPLICANT: Massachusetts Institute of Technology
TITLE OF INVENTION: DNA Encoding the Heme-Regulated
NUMBER OF INVENTION: Eukaryotic Initiation Factor 2 alpha Kinase
CORRESPONDENCE ADDRESSES: 2
ADDRESS: Kilpatrick & Cody
STREET: 1100 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08131
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: MIT 5807
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-815-6508
TELEFAX: 404-815-6555
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2729 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Rabbit
CELL TYPE: Reticulocytes
FEATURE:
NAME/KEY: misc.feature
LOCATION: 113..2149
OTHER INFORMATION: /note= "Expression of HRI mRNA in Human erythroid cells,
FEATURE:
NAME/KEY: misc.feature
LOCATION: 229..249
OTHER INFORMATION: /note= "Primer used in the amplification of human HRI cDN
FEATURE:
NAME/KEY: misc.feature
LOCATION: 543..560
OTHER INFORMATION: /note= "Primer used in the amplification of human HRI cDN
FEATURE:
NAME/KEY: misc.feature
LOCATION: 448..468
OTHER INFORMATION: /note= "Primer used in the amplification of human HRI cDN
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1009..1031
OTHER INFORMATION: /note= "Primer used in the amplification of a human HRI c
PUBLICATION INFORMATION:
AUTHORS: Chen, Jane J.
AUTHORS: London, Irving M.
TITLE: Cloning of the cDNA of the heme-regulated
TITLE: eukaryotic initiation factor 2alpha

TITLE: (eIF-2alpha)kinase of rabbit reticulocytes;
TITLE: Homology to yeast GCN2 protein kinase and human
TITLE: double-stranded-RNA-dependent
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
VOLUME: 88
PAGES: 7729-7733
DATE: September-1991
PCT-US93-08131-1

Query Match 60.0%; Score 15; DB 5; Length 2729;
Best Local Similarity 78.3%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 attcgatcctcgtcaggtgcag 25
Db 2059 ATTCAAGCCTCCTGAGGTTGAG 2081

RESULT 45
US-08-454-097-34/c
Sequence 34, Application US/08454097
Patent No. 5686412
GENERAL INFORMATION:
APPLICANT: Hoechst, Merl F.
TITLE OF INVENTION: Protein Kinases
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESSES:
ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive, 6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,097
FILING DATE: 30-MAY-1995
CLASSIFICATION: 424
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/185,359
FILING DATE: 21-JAN-1994
APPLICATION NUMBER: US 08/008,001
FILING DATE: 21-JAN-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/728,783
FILING DATE: 03-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: No. 5686412and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/31853
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 3505 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 154..1398
US-08-454-097-34

Query Match 60.0%; Score 15; DB 1; Length 3505;

Wed Oct 10 07:44:54 2001

us-09-396-196f-8.std.rn1

Page 22

Best Local Similarity 78.3%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 cattcgatcctcgtcaggtgca 24
|||||
Db 1502 CATTTCGATCCTAGACCGGGGA 1480

Search completed: October 9, 2001, 11:39:59
Job time: 1914 sec